Universidade de Aveiro Departamento de Biologia
2013

João Manuel Salvador Simões

Consequências fenotípicas de erros de tradução em *Candida albicans*

Phenotypic consequences of genome mistranslation in *Candida albicans*

Universidade de Aveiro Departamento de Biologia
2013

João Manuel Salvador Simões

Consequências fenotípicas de erros de tradução em *Candida albicans*

Phenotypic consequences of genome mistranslation in *Candida albicans*

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Biologia, realizada sob a orientação científica do Prof. Doutor Manuel António da Silva Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

o júri

presidente

Doutor Luís António Ferreira Martins Dias Carlos
Professor Catedrático da Universidade de Aveiro

Doutora Isabel Maria Godinho de Sá Nogueira
Professora Associada com Agregação da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Doutor Vítor Manuel Vieira da Costa.
Professor Associado do Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto

Doutora Maria Paula Polónia Gonçalves
Professora Associada da Universidade de Aveiro

Doutor Manuel António da Silva Santos
Professor Associado da Universidade de Aveiro

Doutora Sandra de Macedo Ribeiro
Professora Auxiliar Convidada da Universidade de Aveiro

Doutor Cláudio Emanuel Moreira Gomes
Investigador Auxiliar do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa (ITQB/UNL)

Agradecimentos

Acknowledgements

I would like to thank to my supervisor, Professor Manuel Santos, for the opportunity to work on this PhD project, for all the scientific help and support during the last 5 years and for the guidance and discussions.

I am grateful to all members in the RNA Biology laboratory for the past 5 years of good moments and motivational words. I am especially thankful to Ana Rita Bezerra for help, support and funny moments along this project, to Denisa for the *S. cerevisiae* strains and great help, to Marisa for the constant motivation and for being a cool lab colleague and Violeta for taking care of everything with a never ending patience with all of us. I am thankful to Tatiana, Ana Soares, Cristina and João Paredes for helping me when I started my laboratory work. To Laura, Céu, Patrícia, Ana Rita, Sofia, Gabriela, Tobias, Jörg and to the other past members of the RNA Biology laboratory for their friendship and for making the laboratory a pleasant place to do science.

I am very grateful to Professor Judith Berman from the Universities of Telavive and Minnesota for her help with our reporter system. Professor Teresa Gonçalves and Lisa from the University of Coimbra for helping in the macrophage experiments, and Helena from University of Aveiro for the helping with the flow cytometer experiments.

I am grateful to FCT, for funding my work through a PhD grant SFRH/BD/39146/2007 and project PTDC/BIA-MIC/099826/2008

I am thankful to my house friends Luís, Roberto and Zé for the patient and funny moments.

To my friends that always try to motivate me and cheer me up, André, Bino, Catarina, Inês, Renato and Joel.

Finally, I am grateful to my parents, brother, sister and all family that always support my decisions and especially to my grand-parents for all the good moments.

To all of you: big thanks!

palavras-chave

Candida albicans, código genético, tRNA , erros de tradução, ambiguidade do codão,transcriptoma, estudos fenotípicos.

resumo

Várias espécies do género *Candida* traduzem o codão CUG de leucine como serina. Em *C. albicans* este codão é traduzido pelo $tRNA_{CAG}^{Ser}$ de serina que é reconhecido por leucil- e seril-tRNA sintetases (LeuRS e SerRS), permitindo a incorporação de leucina ou serina em posições com CUG. Em condições padrão de crescimento os codões CUG é incorporam 3% de leucina e 97% de serina, no entanto estes valores são flexíveis uma vez que a incorporação de serina pode variar entre 0.6% e 5% em resposta a condições de stress. Estudos anteriores realizados *in vivo* em *Escherichia coli* sugeriram que a ambiguidade em codões CUG é regulada pela SerRS. De facto, o gene da SerRS de *C. albicans* tem um codão CUG na posição 197 (Ser197) cuja descodificação ambígua resulta na produção de duas isoformas de SerRS. A isoforma SerRS_Leu197 é mais ativa, apesar de menos estável, que a isoforma SerRS_Ser197, suportando a ideia da existência de um *feedback loop* negativo, envolvendo estas duas isoformas de SerRS, a enzima LeuRS e o $tRNA_{CAG}^{Ser}$, que mantem os níveis de incorporação de leucina no codões CUG baixos. Nesta tese demonstramos que tal mecanismo não é operacional nas células de *C. albicans*. De facto, os níveis de incorporação de leucina em codões CUG flutuam drasticamente em resposta a alterações ambientais. Por exemplo, a incorporação de leucina pode chegar a níveis de 49.33% na presença de macrófagos e anfotericina B, mostrando a notória tolerância de *C. albicans* à ambiguidade.

Para compreender a relevância biológica da ambiguidade do código genético em *C. albicans* construímos estirpes que incorporam serina em vários codões. Apesar da taxa crescimento ter sido negativamente afetada em condições padrão de crescimento, as estirpes construídas crescem favoravelmente em várias condições de stresse, sugerindo que a ambiguidade desempenha um papel importante na adaptação a novos nichos ecológicos. O transcriptoma das estirpes construídas de *C. albicans* e *Saccharomyces. cerevisiae* mostram que as leveduras respondem à ambiguidade dos codões de modo distinto. A ambiguidade induziu uma desregulação moderada da expressão génica de *C. albicans*, mas ativou uma resposta comum ao stresse em *S. cerevisiae*. O único processo celular que foi induzido na maioria das estirpes foi a oxidação redução. De salientar, que enriquecimento em elementos cis de fatores de transcrição que regulam a resposta à ambiguidade em ambas as leveduras foi distinta, sugerindo que ambas respondem ao stresse de modo diferente.

Na globalidade, o nosso estudo aprofunda o conhecimento da elevada tolerância à ambiguidade de codões em *C. albicans*. Os resultados sugerem que este fungo usa a ambiguidade do codão CUG durante infeção, possivelmente para modular a sua interação com o hospedeiro e a resposta a drogas antifúngicas.

keywords

Candida albicans, genetic code, tRNA, mistranslation, codon ambiguity, transcriptome, phenomics.

abstract

Several *Candida* species are characterized by the reassignment of the leucine CUG codon to serine. In *C. albicans* this codon is translated by a serine ($tRNA_{CAG}^{Ser}$) that is recognized by both the leucyl - and the seryl-tRNA synthetases (LeuRS and SerRS), allowing for incorporation of leucine and serine at CUG positions. Under standard growth conditions 3% of leucine and 97% of serine are incorporated at CUGs, but these values are flexible as leucine misincorporation fluctuates between 0.6% and 5% in response to stress conditions. Previous *in vivo* studies carried out in the heterologous host *Escherichia coli* suggested that CUG ambiguity is regulated by the SerRS. Indeed, the *C. albicans* SerRS gene has one CUG codon at position 197 (Ser197) whose ambiguous decoding results in the production of two SerRS isoforms. The SerRS_Leu197 isoform is more active, although less stable, than the SerRs_Ser197 isoform, supporting the existence of a negative feedback loop involving these two isoforms of the SerRS, the LeuRS enzyme and the $tRNA_{CAG}^{Ser}$ that maintain low levels of leucine incorporation at CUG codons. We show in this thesis that such mechanism is not operational in *C. albicans* cells. Indeed, the *in vivo* levels of Leucine incorporation at CUGs fluctuate rather dramatically in response to environmental cues. For example, Leucine incorporation is as high as 49.33% in presence of macrophages and amphotericin B, showing remarkable tolerance of *C. albicans* to CUG ambiguity.

To understand the biological relevance of genetic code ambiguity in *C. albicans* we have engineered strains that misincorporate Serine at various non-cognate codons. Although growth rate was negatively affected in normal growth conditions, the engineered cells displayed growth advantages in several stress conditions, suggesting that ambiguity plays an important role in adaptation to new ecological niches. Transcriptome profiling of engineered *C. albicans* and *Saccharomyces cerevisiae* strains showed that both yeasts respond to codon ambiguity in remarkably different ways. Codon ambiguity induced mild gene expression deregulation in *C. albicans*, but activated a general common stress response in *S. cerevisiae*. The only cellular process that was induced in almost all strains was oxidation reduction. More importantly, the transcription factors that regulate the response to codon ambiguity in both yeasts were distinct, suggesting that both yeasts respond to stress in fundamentally different ways.

The overall study provides deep insight on the very high level of tolerance to codon ambiguity of the main human pathogen *C. albicans*. The data strongly suggest that this fungus uses CUG ambiguity during infection, likely to modulate host-pathogen and antifungal drug responses.

List of contents

List of figures

List of tables

List of abbreviations

aaRS	aminoacyl-tRNA synthetase
aa-tRNA	aminoacyl-tRNA
ATP	adenosine 5'-triphosphate
bp	Base pair
cDNA	complementary deoxyribonucleic acid
Cy3	Cyanine 3
$CuSO_4$	copper sulfate
DNA	deoxyribonucleic acid
DNase	deoxyribonulease
dNTP	deoxynucleotide solution mix
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
ESR	environmental stress response
g (mg, µg, ng)	gram (milligram, microgram, nanogram)
GO	gene ontology
HSP	heat-shock protein
L (mL, µL)	liter (mililiter, microliter)
LiCl	lithium chloride
LOH	loss of heterozygosity
M (mM, µM)	molar (milimolar, micromolar)
mQ	milliQ
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NADH	nicotinamide adenine dinucleotide

$O.D_{600}$	optical density at 600nm
ORF	open reading frame
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
RF1	release factor-1
RNA	ribonucleic acid
RNase	ribonuclease
ROS	reactive oxygen species
rpm	revolutions per minute
RPS10	ribosomal protein S10
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
V	volt
yEGFP	yeast-enhanced green fluorescent protein
YPD	yeast extract peptone dextrose

Other abbreviations will be explained when used in the text.

1. Introduction

1.1. Genetic Code

Crick defined the genetic code as *a non-overlapping triplet code. Most, but not all, of the 64 triplets stand for one or another of the 20 amino acids and, in most cases, each amino acid is represented by more than one codon* (Crick, 1968). This definition of the genetic code summarized in Table 1-1 represents a modest subset of the complexity of translation encoded by an organism's DNA, and even smaller than transcription and other molecular interactions (Berleant *et al.*, 2009). In 1961, Nirenberg and Matthaei discovered that *Escherichia coli* extracts programmed with poly(U) in solution with radioactive amino acid phenylalanine resulted in the incorporation of only phenylalanine into protein (Nirenberg and Matthaei, 1961). Interaction experiments between doublets or triplets consisting only of U residues with 3H-phenylalanine tRNA and ribosomes resulted in binding when using the triplet UUU but no effect was observed with the doublet UU, indicating that the genetic code is organized in triplets (Nirenberg and Leder, 1964).

Table 1-1: Standard genetic code table.

First letter	Second letter: U		Second letter: C		Second letter: A		Second letter: G		Third letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Although the standard genetic code is composed by 64 codons, not all codons correspond to amino acids. For example, UAA, UAG and UGA codons correspond to termination of protein synthesis. The AUG codon can be recognized by N-formyl-methionine-tRNA to initiate protein synthesis and at internal positions corresponds to methionine residues (Nirenberg, 2004). The standard genetic code table shows that codons with U as a second base correspond to hydrophobic amino acids like phenylalanine (Phe), leucine (Leu), isoleucine (Ile), methionine (Met) and valine (Val). The codons with A as the second base correspond to hydrophilic amino acids. And amino acids with chemically similar side chains as aspartic acid (Asp), glutamic acid (Glu), asparagine (Asn) and glutamine (Gln) have chemically similar codons (Nirenberg, 2004). Even if the codon table assigns different codons to the same amino acid, these synonymous codons are not equivalent. In fact, synonymous codons appear in different proportions in different species, these codon usage biases are linked with gene expression levels, translation efficiency and protein folding (Gu *et al.*, 2004;Angellotti *et al.*, 2007;Berleant *et al.*, 2009).

1.2. Origin and evolution of the Genetic Code

When the genetic code was discovered it was postulated that it would be the same in all organisms (universal code), as any change would be lethal or very strongly selected against. Since alterations in the genetic code would change the meaning of codons, leading to erroneous translation and subsequently to disruption of proteins, it was assumed that the genetic code could not evolve. The allocation of codons to amino acids was also viewed as a matter of chance, a result of a “frozen accident” (Crick, 1968). Several theories emerged to counteract this “Frozen accident theory”, of the origin and evolution of the genetic code.

1.2.1. Stereochemical theory

This theory postulates that the codon-to-amino acids assignments is determined by physicochemical affinity, between the amino acids and the cognate nucleotide triplets (codons or anticodons). Thus the specific structure of the code is not accidental but rather necessary and possibly unique (Di, 2005;Koonin and Novozhilov, 2009) . A similar hypothesis known as the "escaped triplet theory", claims that the primordial code was based on interactions between amino acids and cognate triplets located within amino acid binding pockets of RNA molecules (Yarus *et al.*, 2005; Koonin and Novozhilov, 2009). The main observation that sustains this theory is that RNA aptamers, selected from random sequences, bind significantly to amino acids corresponding to cognate triples (codons or anticodons) of the standard genetic code (Koonin and Novozhilov, 2009). From the eight observed amino acids, namely arginine, glutamine, histidine, isoleucine, leucine, phenylalanine, tryptophan and tyrosine, only glutamine revealed no significant correlation with codons or anticodons in aptamer selection experiments. Conversely, arginine binds strongly to RNA aptamers enriched in arginine codons (Yarus *et al.*, 2005). When the standard genetic code was compared with random alternatives codes, only 9.7% of random codes revealed greater codon association, and 0.2% greater anticodon association with selected aptamers. Although the aptamers experiment strongly supports the stereochemical theory, some points should be highlighted. Firstly the statistical models used may be questionable and secondly, the aptamer experiments and the interaction with codon, anticodon and amino acids show a lack of coherence that is difficult to link with the physical basis of the code (Koonin and Novozhilov, 2009).

1.2.2. Adaptive theory

The adaptive theory asserts that the genetic code structure was shaped under selective forces, leading to maximum robustness and minimizing the effect of errors on proteins structure and function (Koonin and Novozhilov, 2009). This theory is supported by two main hypothesis, namely the "lethal-mutation"

hypothesis, which claims that the standard code evolved to minimize the consequence of point mutations (Koonin and Novozhilov, 2009) and the "translation error minimization" hypothesis which states that the code evolved through a force that minimizes effects of translational reading errors (Woese, 1965).

Several observations sustain this theory. Firstly, related codons tend to code for similar amino acids. Secondly, mistranslation occurs in the first and third position of codons with higher incidence, and it is very rare in the second codon position where there is stronger correlation with amino acids properties (Woese, 1965).

To better support the "translation error minimization" hypothesis, quantitative evidences were obtained by comparing the standard genetic code with random alternative codes. Freeland and Hurst calculated the average effect of shifting a codon by a single base for all possible single base shifts in the genetic code and for changes in the three codon positions separately (Haig and Hurst, 1991). Such values were obtained for amino acids attributes, namely polar requirement, hydropathy, molecular volume and isoelectric point (Haig and Hurst, 1991). The natural genetic code showed better results when compared with the random codes except for 0.02% of the time (Haig and Hurst, 1991).

Therefore, although the exact selective forces responsible for the structure of the standard genetic code are difficult to identify, the genetic code is robust to translational misreading and mutation (Koonin and Novozhilov, 2009).

1.2.3. Coevolution theory

The coevolution theory asserts that the organization of the standard genetic code reflects the pathways of amino acid biosynthesis (Wong, 2005;Koonin and Novozhilov, 2009). This theory proposes that prebiotic synthesis would not be able to produce the 20 proteinaceous amino acids. Therefore, several amino acids had to be biosynthesized before being introduced into the genetic code and translation, leading to a coevolution of the code and amino acid metabolism (Koonin and Novozhilov, 2009). This premise is based on the observed organization of the

genetic code. Firstly the Asp-family of amino acids, which have Asp as biosynthetic precursor, occupy three codon boxes across the ANN row of the code, and Ile and Met sharing the AUN codon box are derived from Asp through homoserine. In the UGN codon box are Cys and Trp, which are biosynthetic products of Ser (Wong, 2007). Consequently, codon assignment to amino acids could have been driven by metabolic connections between the amino acids (Koonin and Novozhilov, 2009).

Under this theory, there were three main phases of incorporation of amino acids into the genetic code. The first phase is characterized by inclusion of amino acids from prebiotic synthesis, the second phase is characterized by the introduction of amino acids biosynthesized from first phase amino acids, and finally in the third phase amino acids were introduced into proteins through posttranslational modifications (Koonin and Novozhilov, 2009).

The selection of amino acids of the first phase is supported by the highest yield of amino acids obtained in Miller´s mix, namely Ala, Gly, Asp and Val (Figure 1-1) (Trifonov, 2004). Incorporation of new amino acids in the genetic code, produced by the evolution of metabolic pathways, having in account precursor-product pair of amino acids is illustrated in Figure 1-1. The demonstration that amino acid composition of proteins is evolvable also supports this theory. For example, *Bacillus subtilis* could be engineered to substitute tryptophan by 4-fluorotryptophan and even to replace tryptophan (Wong, 1983).

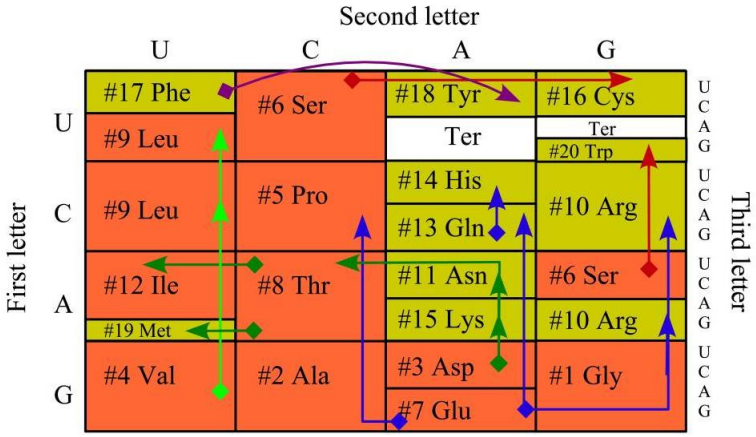


Figure 1-1: Standard genetic code expansion according to the coevolution theory. First phase amino acids are orange, second phase amino acids are green. Numbers indicate the proposed order of amino acid appearance in the code. Arrows define 13 precursor-product pairs of amino acids, their colour identify the biosynthetic families of Glu (blue), Asp (dark-green), Phe (magenta), Ser (red), and Val (light-green). Adapted from (Koonin and Novozhilov, 2009)

Two main criticisms can be pointed to the coevolution theory. Firstly, the coevolution theory is highly dependent on the selection of amino acid precursor-product pairs, and this selection is far from being simple. When Wong proposed the coevolution theory, he used inferred reactions of primordial metabolism between amino acids that are questionable, instead of using biochemically established relationships between amino acids (Koonin and Novozhilov, 2009). A second criticism is the biological validity of the statistical analysis of Wong (Ronneberg *et al.*, 2000) and the invalid assumption that no other process might group amino acids into NNY codon blocks. In fact, the probability of a randomly generated code displaying a biosynthetic pattern as strong or stronger than those showed by the canonical code is 23%. This probability would increase to 62%, if the assumption of the chronological order in which codons were reassigned from precursor to product amino acid are taken in account (Ronneberg *et al.*, 2000).

1.3. Translation

1.3.1. Transfer RNA

1.3.1.1. Structure

Transfer RNAs (tRNA) (Figure 1-2) are normally composed by 76 nucleotides folded into a cloverleaf structure. This structure is formed by an acceptor stem (amino acid attachment site), a D-stem loop (D since dihydrouridine is present and conserved), a TΨC-stem-loop (T because ribothymidine is present and conserved), an anticodon stem-loop (anticodon triplet at nucleotide in position 34, 35 and 36) and a variable region (Giege, 2008).

The existence of conserved and semi-conserved residues molds the tRNA cloverleaf into a three dimensional L-shaped structure with two domains (Figure 1-2) (Ribas de Pouplana and Schimmel, 2001a;Giege, 2008). One of these domains is composed by the acceptor stem and T-stem loop, linked to the other domain, which is composed by the D-stem-loop and anticodon stem-loop. Consequently, the anticodon and amino acid attachment site are in different domains separated by 75 Å (Ribas de Pouplana and Schimmel, 2001a).

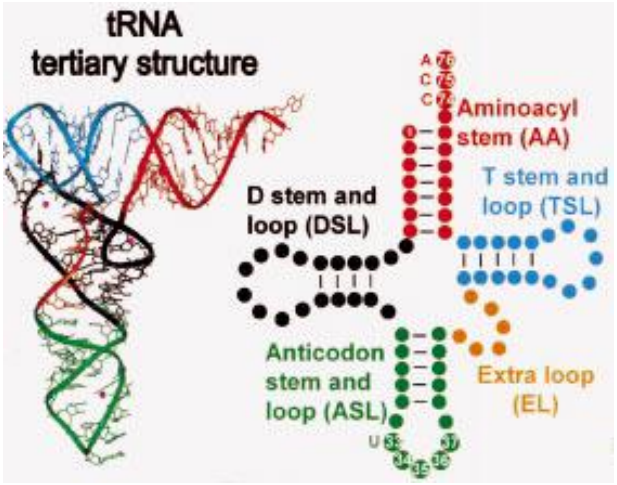


Figure 1-2: tRNA 3D and 2D structures. tRNA tertiary structure with the L-shaped (left image) and 2D tRNA with cloverleaf structure (right image). In both structures colours correspond to: Red) Aminoacyl stem; Blue) T stem; Yellow) extra loop; Green) anticodon stem; Black) D stem. Adapted from (Agris, 2004)

The cloverleaf form shows variation. In mitochondrial tRNAs entire domains are frequently missing and in virus tRNA-like structures the acceptor stem is replaced by a pseudoknot. The L-shaped structure can assume conformational changes depending on the functional state of the tRNA, which can be free or in a complex with macromolecular ligands, resulting in large variations in the angle between the two branches and acceptor stem (Giege, 2008).

1.3.1.2. tRNA processing

tRNA genes have intragenic promoters containing two central elements, named A and B boxes. These promoters are conserved in tRNA genes because they encode the tRNA D- and TΨC-loops. The space between these two boxes is variable depending on the presence of introns in the anticodon-loop domain. The transcription of tRNA genes in eukaryotes is carried out by RNA polymerase III, being the A and B boxes essential for binding of transcription factors and subsequent transcription initiation (Schramm and Hernandez, 2002;Marck *et al.*, 2006).

The tRNA transcript molecule undergoes several processing reactions before becoming a mature and functional tRNA (Figure 1-3). These processes

involve removal of the 5’ and 3’ end trailer sequences, addition of 3’ terminal CCA sequence, removal of introns and several modifications of tRNA nucleotides. The 5’ trailer is removed by ribonuclease P (RNase P). This ribozyme is universally conserved and generates a tRNA molecule containing a mature 5’-end with a phosphate, releasing simultaneously a sequence with a 3’-OH terminus (Nakanishi and Nureki, 2005;Michelle Mitchell and Hong Li, 2009;Phizicky and Hopper, 2010). Maturation of the tRNA 3’ terminus is processed by several enzymes, namely the endonucleases RNase E and RNase III, which cleave the 3’ trailer sequence in the middle. Subsequently, the residual trailer is cleaved by the exonucleases RNase II, RNase BN and RNase PH. Finally, the endonuclease RNase Z cleaves the precursor tRNA immediately after the discriminator base, resulting in a free 3’ hydroxyl prepared for CCA addition. The CCA sequence synthesis or reparation is accomplished by a template-independent nucleotidyltransferase, the CCA-adding enzyme, using CTP and ATP as substrates. This process starts with the CCA-adding enzyme unfolding the tRNA helix composed by the acceptor and TΨC-arm, measuring the RNA helix, and selecting only tRNAs as substrate. This enzyme has a L-shaped catalytic pocket, where the terminal of the tRNA is inserted with stacking of the nucleotides, which are fixed by complementary pockets formed by the protein amino acid residues that mimic the DNA template and subsequently resulting in CCA synthesis (Nakanishi and Nureki, 2005;Michelle Mitchell and Hong Li, 2009).

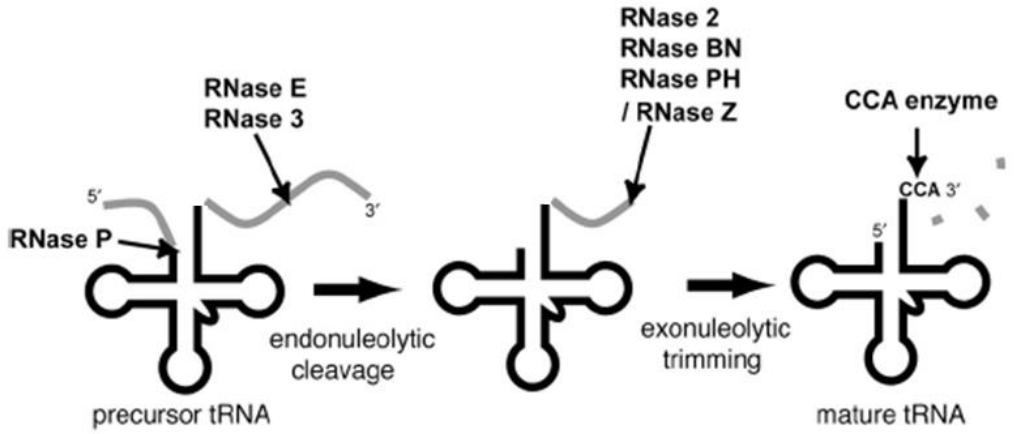


Figure 1-3: The tRNA processing of the 5′ leader sequence and 3′ trailer sequence. The 5’leader sequence is excised by RNase P, while 3’trailer is endonucleolytic cleaved by RNase E and RNase 3, and exonucleolytic trimmed by RNase 2, RNase BN, RNase PH and RNase Z. Finally the CCA is added to the 3’ trailer by a nucleotidyltransferase. Adapted (Nakanishi and Nureki, 2005)

Several tRNA genes have introns. In yeast 20% of the tRNA genes contain introns. These can vary between 14 to 60 nucleotides and interrupt the anticodon loop immediately 3' to the anticodon. Although intron sequence is not conserved at the splice junctions, the 3' splice junction is in a bulged loop. The location of introns and their characteristics are generally conserved, but in archaea the location can differ. Removal of the introns from pre-tRNA in organelles and in bacteria is accomplished by a self-splicing reaction, however in eukaryotes and archaea the removal is catalyzed by proteins (Michelle Mitchell and Hong Li, 2009;Phizicky and Hopper, 2010). In budding yeasts this pre-tRNA splicing reaction engages several steps, being each step catalyzed by a different enzyme (Figure 1-4). Firstly, the pre-tRNA is cleaved at two splice sites by a heterotetrameric endonuclease (constituted of Sen2, Sen15, Sen34, and Sen54 in yeast) resulting in 5' half molecules with a 2'–3' cyclic phosphate and 3' half molecules with a 5' hydroxyl group. In the second step, a monomeric tRNA splicing ligase (Trl1) promotes the ligation of the half molecules by adding a phosphate to the 3´-tRNA half and binding the 5' and 3' halves through a phosphodiester bond. In the last step, the tRNA 2' phosphotransferase (Tpt1) removes the extra 2' phosphate at the splice junction. Subsequently, the 2'phosphate is transferred by Tpt1 to form ADP-ribose 1"–2" cyclic phosphate (Abelson *et al.*, 1998;Hopper *et al.*, 2010;Phizicky and Hopper, 2010).

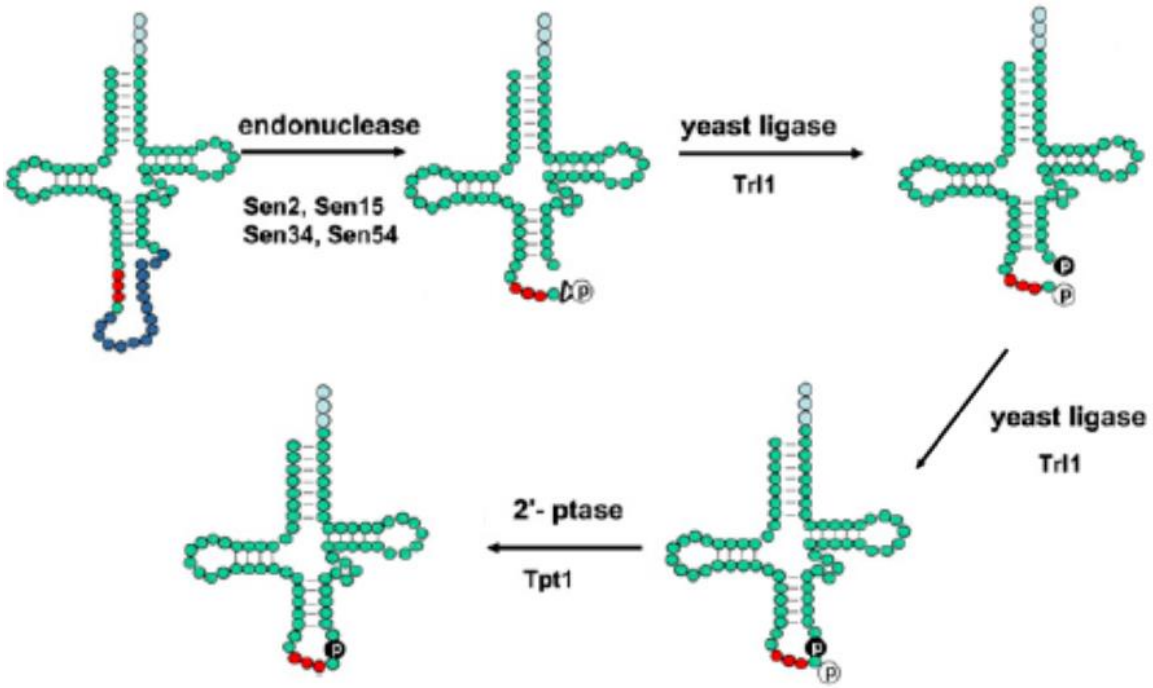


Figure 1-4: tRNA splicing and ligation pathways in yeast. tRNA is displayed in secondary structure with the anticodon indicated by red circles and the intron indicated by blue circles. The pre-tRNA is cleaved by an endonuclease, resulting in 5' half molecules with a cyclic phosphate (triangle with a white circle containing P) and a 5'-OH. Next, the ligase Trl1 promotes the ligation of the half molecules by adding a phosphate to the 3´-tRNA (black circle containing the P) half and binding the 5' and 3' halves by a phosphodiester bond. Finally a tRNA 2'phosphotransferase (Tpt1) removes the extra phosphate (white circle containing P). Adapted from (Phizicky and Hopper, 2010).

1.3.1.3. tRNA modification

RNA nucleotides are frequently modified. Indeed, more than one hundred nucleotide modifications have been described in RNA, the majority of them in tRNAs (http://library.med.utah.edu/RNAmods/) (Czerwoniec *et al.*, 2009). RNA modifications can vary from methylation of all four nucleotides bases or sugar, uracil isomerization to pseudouracil or conversion to dihydrouracil, cytidine N-acetylation and adenosine conversion to inosine. Modifications occurring at the anticodon normally involve guanosine conversion to wybutosine or queuosine, adenosine isopentylation and uracil methoxycarbonylmethylation or thiolylation (Figure 1-5) (Hopper *et al.*, 2010).

tRNA modifications are split into two main categories. Modifications that contribute to stabilize the tRNA L-shaped tertiary structure that are localized in the

tRNA core region (D- and TΨC loop). Also there are modifications localized in the anticodon loop that are involved with codon pairing precision and aminoacyl-tRNA synthetases recognition (Nakanishi and Nureki, 2005).

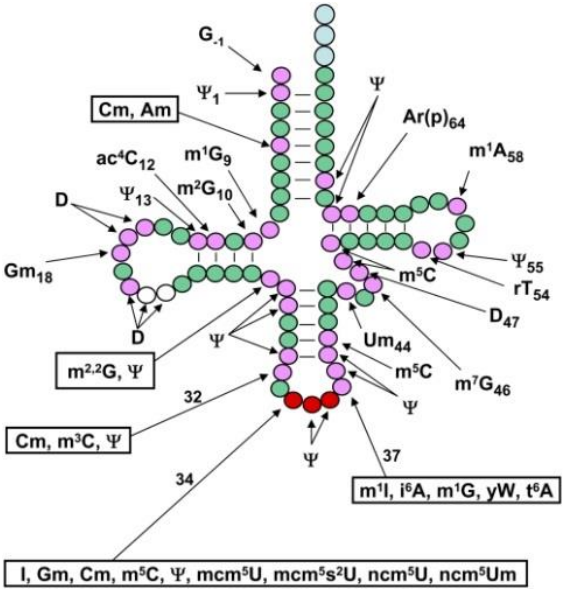


Figure 1-5: Scheme of the location of modifications in cytoplasmic tRNA in *S. cerevisiae*. tRNA is displayed in secondary structure. The tRNAs residues number may vary, however the anticodon is always numbered as residues 34, 35, and 36, and the CCA end is always numbered as 74, 75, and 76. Green circles indicate residues that are unmodified, pink circles represent residues that may be modified, white circles represent residues 20a and 20b that could be present in some tRNA and are sometimes modified, light-blue circles indicate the CCA end and red circles indicate the anticodon that are sometimes modified. The conventional abbreviation used is described in the Modomics database (http://modomics.genesilico.pl). *(Ψ) Pseudouridine, (Am) 2'-O-methyladenosine, (Cm) 2'-O-methylcytidine, (m^1G) 1-methylguanosine, (m^2G) 2-methylguanosine, (ac^4C) 4-acetylcytidine, (D) dihydrouridine, (Gm) 2'-O-methylguanosine, ($m^{2,2}G$) N2,N2-dimethylguanosine, (m^3C) 3-methylcytidine, (I) inosine; (m^5C) 5-methylcytidine, (mcm^5U) 5-methoxycarbonylmethyluridine, (mcm^5s^2U) 5-methoxycarbonylmethyl-2-thiouridine, (ncm^5U) 5-carbamoylmethyluridine, (ncm^5Um) 5-carbamoylmethyl-2'-O-methyluridine, (m^1I) 1-methylinosine, (i^6A) N6-isopentenyl adenosine, (yW) wybutosine, (t^6A) N6-threonylcarbamoyladenosine, (Um) 2'-O-methyluridine, (m^7G) 7-methylguanosine, (rT) ribothymidine, [Ar(p)] 2'-O-ribosyladenosine (phosphate).* Adapted from (Phizicky and Hopper, 2010).

1.3.1.3. tRNA turnover

The tRNA integrity is guaranteed by several surveillance mechanisms. The first occurs at the pre-tRNA level by the TRAMP complex. A second surveillance step is carried out at the nuclear export level by tRNA synthetases located in the

nucleus. A third surveillance step takes place at mature tRNA level through rapid tRNA decay (RTD), targeting hypomodified tRNAs (Chernyakov *et al.*, 2008).

The TRAMP complex and exosome turnover pathway serves as a nuclear supervision to guarantee that only functional correctly structured tRNAs are delivered to the protein synthesis machinery in the cytoplasm. The TRAMP complex includes a poly(A) polymerase (Trf4 or Trf5), a zinc-knuckle putative RNA-binding protein (Air1 or Air2), and a RNA helicase (Mtr4). It binds to defective tRNAs and adds a short poly(A) tail, targeting them for degradation. Additionally TRAMP recruits the exosome responsible for tRNA degradation (Houseley and Tollervey, 2009;Hopper *et al.*, 2010;Phizicky and Hopper, 2010).

The rapid tRNA decay (RTD) pathway degrades hypomodified tRNAs and is independent from the TRAMP complex and nuclear exosome. This pathway was discovered by the fast turnover of tRNAs with certain tRNA modifications. The components involved in the RTD pathway are Met22 and 5'-3' exonucleases Rat1 and Xrn1, Met22 acts indirectly through Rat1 and Xrn1. This degradation process is proposed to occur in the nucleus and cytoplasm, since Rat1 is located in the nucleus and Xrn1 in the cytoplasm. An example of this process is the mature $tRNA_{AAC}^{Val}$ lacking 7-methylguanosine and 5-methylcytidine which is rapidly degraded and deacylated at 37°C in a strain with *trm8* and *trm4* genes deleted resulting in temperature-sensitive growth. Deletion of Met22 or mutation of Rat1 and Xrn1 genes in this strain prevented the degradation and deacylation of $tRNA_{AAC}^{Val}$ restoring growth to normal levels at 37ºC (Chernyakov *et al.*, 2008).

The endonucleolytic cleavage of tRNA in the anticodon loop, producing tRNA halves, is catalyzed by endonucleases in the cytoplasm, namely Rny1p (Hopper *et al.*, 2010). This process has been shown in yeast under oxidative stress, stressed mammalian cells, protozoa and bacteria under nutrient starvation. In yeast, the enzyme responsible for tRNA degradation is Rny1p, which belongs to the RNase T2 family, which is normally located in the vacuole. In other words, under oxidative stress, the cytosolic tRNA is cleaved by Rny1p which is released from the vacuole into the cytosol (Thompson and Parker, 2009).

1.3.2. Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases (aaRSs) are the enzymes responsible for catalyzing the covalent attachment of amino acids to the 3' end of tRNAs. This reaction is denominated aminoacylation and its accuracy is mainly affected by amino acids chemical and physical similarity and tRNAs structural similarity. Despite this, aminoacylation error rate is in the order of 10^{-6} in the two distinct half reactions (Figure 1-6). In the first step of aminoacylation the amino acid is activated by condensation with ATP, forming the enzyme-bound aminoacyl-adenylate intermediate with release of pyrophosphate (PPi). In the second step, this intermediate is attacked by the 2′-OH (Class I) or 3′-OH (Class II) of the tRNA's 3' end, producing aminoacyl-tRNA and releasing AMP and ARS (Francklyn, 2008).

$$\mathrm{aa+ATP+ARS} \xleftrightarrow{\mathrm{Mg2+}} \mathrm{ARS{:}(aa-AMP)+PPi} \quad (1)$$

$$\mathrm{ARS{:}(aa-AMP)+tRNA} \leftrightarrow \mathrm{aa-tRNA+AMP+ARS} \quad (2)$$

Figure 1-6: The aminoacyl-tRNA synthetase reaction mechanism. (1) The aminoacyl-tRNA synthetase (ARS) binds the amino acid (aa) and through a reaction with ATP and Mg^{2+} forms the aminoacyl-adenylate (aa-AMP) with release of a pyrophosphate molecule (PPi). (2) The aa-AMP reacts with the tRNA Amino acid transfer and the aa is transferred to the tRNA (aa-tRNA), resulting in the tRNA esterification and release of AMP. Adapted from (Francklyn, 2008).

1.3.2.1. Classes and subclasses of aminoacyl-tRNA synthetases

The aaRSs are distributed into two distinct classes (Figure 1-7). The class I is composed by ArgRS, CysRS, GluRS, GlnRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS and ValRS. These enzymes are characterized by an amino-terminal Rossmann fold domain with "KMSKS" and "HIGH" signature sequences. The Rossmann fold domain is responsible for binding the amino acid and ATP, activate the amino acid and transfer it to the 3' end of tRNA. The carboxyl-terminal domain is responsible for the tRNA anticodon recognition, however the domain structures

is highly variable among class I aminoacyl-tRNA synthetases, indicating a possible late appearance of the Rossmann fold domain. The acceptor stem of the tRNA is recognized by the connective peptide (CP1) insertion domain, which divides the Rossmann fold domain. The attachment of the amino acid is made with the 2' OH tRNA in the class I aminoacyl-tRNA synthetases (Cavarelli and Moras, 1993).

The class I aaRSs are divided into three subclasses designated Ia, Ib and Ic. All members of subclass Ia and Ib are monomers, while subclass Ic members are homodimers. Enzymes in each subclass recognize chemically related amino acids. Subclass Ia enzymes identify Arg, hydrophobic amino acids (Ile, Leu and Val) and the sulfur containing amino acids Met and Cys. Subclass Ib enzymes are responsible for recognition of the charged amino acids Glu and Lys and the derivative Gln, while subclass Ic enzymes recognize the aromatic amino acids Tyr and Trp (Arnez and Moras, 1997;Ribas de Pouplana and Schimmel, 2001a).

The class II includes AlaRS, AsnRS, AspRS, GlyRS, HisRS, PheRS, ProRS, SerRS, ThrRS and normally LysRS. This class is characterized by a catalytic core composed of seven-stranded β-sheets flanked by three α-helices. Generally, class II aminoacyl-tRNA synthetases are homodimers, except phenylalanyl-, alanyl-, and eubacterial glycyl-tRNA synthetases (Hieronim Jakubowski, 2005). The active site of this class of enzymes consists of a large cavity located on the seven-stranded β-sheet side. This class of enzymes shows three conserved motifs. Motif 1 is part of the dimer interface and assists the orientation of the motifs 2 and 3, which are located near the active site of the enzyme. Motif 2 is involved in the binding of ATP, amino acid and the end of the tRNA, and the motif 3 interacts with the ATP substrate. Generally, class II aminoacyl-tRNA synthetases bind the amino acid to the 3' OH of the ribose of the tRNA terminal adenosine. The only exception is the phenylalanyl-tRNA synthetase, which aminoacylates the 2' OH, but, the amino acid is subsequently transferred to the 3' OH by EF-Tu (Hieronim Jakubowski, 2005).

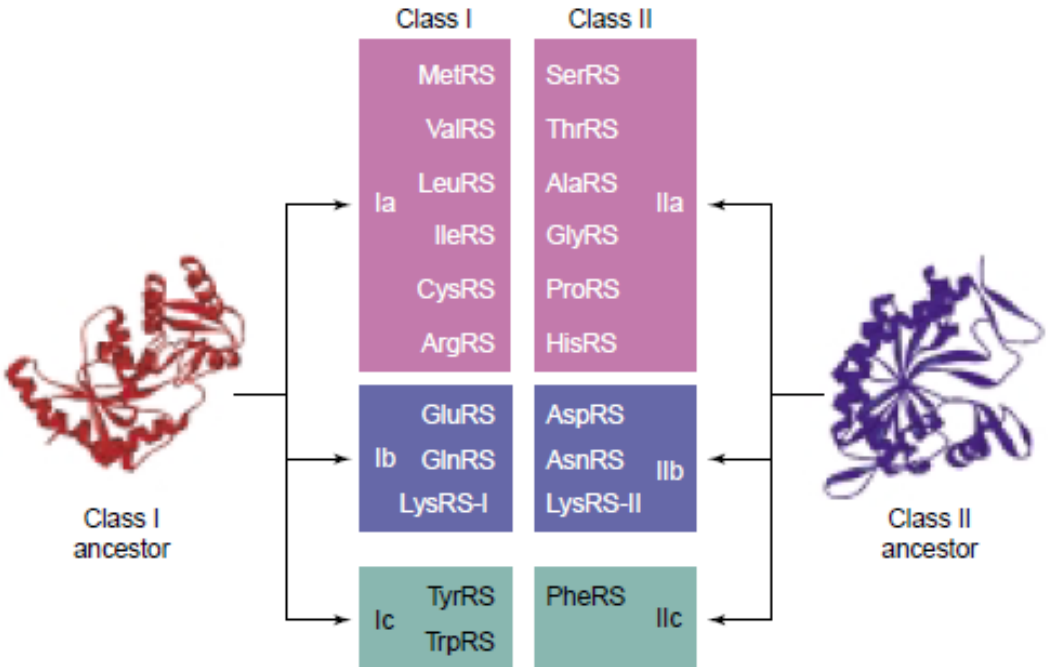


Figure 1-7: Aminoacyl-tRNA synthetases, class and subclass distribution. The aaRSs are distributed in two major classes and subsequently divided in 3 subclasses. Adapted from (Ribas de Pouplana and Schimmel, 2001a).

The class II aminoacyl-tRNA synthetases are also divided in three subclasses IIa, IIb and IIc. As before this division is based on amino acid sequences, structure, location and topology of anticodon binding domains. The subclass IIa enzymes recognize Ser, Thr, His, Gly, Ala and Pro. Generally, the anticodon binding domain of these enzymes is located in the carboxy terminus side of the catalytic domain; the only exceptions are the alanyl- and seryl-tRNA synthetases that do not have an anticodon binding domain. This observation is corroborated by the fact that these enzymes do not identify the anticodon loops of their cognate tRNAs. The subclass IIb enzymes recognize the charged amino acids Asp and Lys and the derivate Asn. Their anticodon binding domains are located on the amino-terminal side of the catalytic domain. Finally, subclass IIc enzymes recognize the aromatic amino acid Phe and its anticodon binding domain is located apart from the catalytic domain (Arnez and Moras, 1997;Ribas de Pouplana and Schimmel, 2001a;Hieronim Jakubowski, 2005)

Comparison of the aaRSs subclasses showed that subclasses Ib and IIb recognize charged amino acids and their derivatives, and subclasses Ic and IIc recognize aromatic amino acids. This symmetry between the two classes is also

observed from a different perspective, namely in the way they approach the acceptor stem. While members of class I approach the acceptor stem from the minor groove side, the class II aaRSs approach it from the major groove side. Since the two classes bind to opposite sides of the tRNA acceptor stem, the possibility of simultaneous binding from each class arises. This was investigated by molecular modeling approaches (Ribas de Pouplana and Schimmel, 2001b) which revealed that subclass-specific pairing was possible without steric clashes, although some combinations revealed steric clashes between the enzymes. The best combinations were between the corresponding subclasses of each class, specifically, Ia with IIa, Ib with IIb and finally Ic with IIc. The best example is between TyrRS and PheRS where modeling revealed acceptor stem mutual accommodation (Ribas de Pouplana and Schimmel, 2001a). These results support the idea that in the primordial genetic code aaRS complementary pairs protected tRNA from the environment (high temperature) by covering the acceptor stem. During this early stage of the code development the charging reactions may have been carried out by ribozymes, which were subsequently replaced by aaRSs. Posterior duplications and mutation of tRNA genes would have led to the final genetic code. This process was associated with changes and duplications of the aaRS genes. These results support the hypothesis that early genomes used both strands of the DNA duplex to code for aaRSs, producing proteins with complementary sequences (Schimmel, 2008).

1.3.2.2. aaRS editing activities

The charging of tRNAs with amino acids is a crucial step in protein synthesis where errors must be prevented to avoid synthesis of aberrant proteins. Aminoacyl-tRNA synthesis requires accurate selection of amino acids and discriminator capacity to select the correct tRNA for charging. To prevent the incorrect charging of near-cognate amino acids, aaRSs have several quality controls.

For some aaRSs the complete aminoacylation reaction is a two-step reaction where the amino acid is activated in the first step (Figure 1-8). This is an

inaccurate step and is followed by a second step where noncognate amino acids are removed. This second step is denominated proofreading or editing (Guo and Schimmel, 2012).

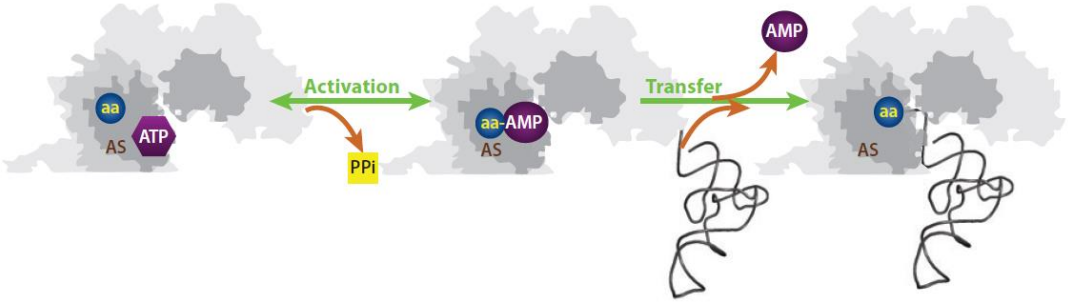


Figure 1-8: The aminoacylation reaction. First the amino acid (aa) is activated at the aminoacyl-tRNA synthetase (aaRS) active site (AS) with ATP forming aminoacyl-adenylate (aa-AMP) resulting in the release of pyrophosphate (PPi), then it is transferred to the 3′ end of the tRNA. Adapted from (Ling *et al.*, 2009).

Half of the 20 canonical aaRSs are absolutely selective. The remaining AARSs, namely AlaRS, MetRS, IleRS, LeuRS, LysRS, GlnRS, ProRS, PheRS, ThrRS and ValRS are less selective and generate incorrect AA-AMPs at rates of 100 to 1000 slower relative to those of the correct AA-AMP (Jakubowski, 2011). To maintain translation fidelity the second group of aaRSs display pre- and post-transfer editing mechanisms (Figure 1-9) (Ling *et al.*, 2009).

The pre-transfer editing can be tRNA independent or tRNA dependent. It occurs before the activated amino acid is transferred to the tRNA 3' end and acts through several mechanisms that promote non-cognate aminoacyl adenylate hydrolysis, resulting in the release of the amino acid, AMP and PPi (Ling *et al.*, 2009;Reynolds *et al.*, 2010).

The post-transfer editing of noncognate aminoacyl-tRNA requires that the amino acid linked to the tRNA CCA 3' end moves from the synthetic site to the editing site, where the tRNA-amino acid ester link is hydrolysed. If editing is processed by the aaRS that synthesized the near-cognate product it is called *Cis*-editing. If editing is processed after release of the tRNA from the active site (after aminoacylation) and is performed by the aaRS or another *trans*-acting factor it is called *trans*-editing. The trans-acting factors can be d-aminoacyl tRNA

deacylases, namely d-Tyr-tRNA deacylase which is widespread (Reynolds *et al.*, 2010)

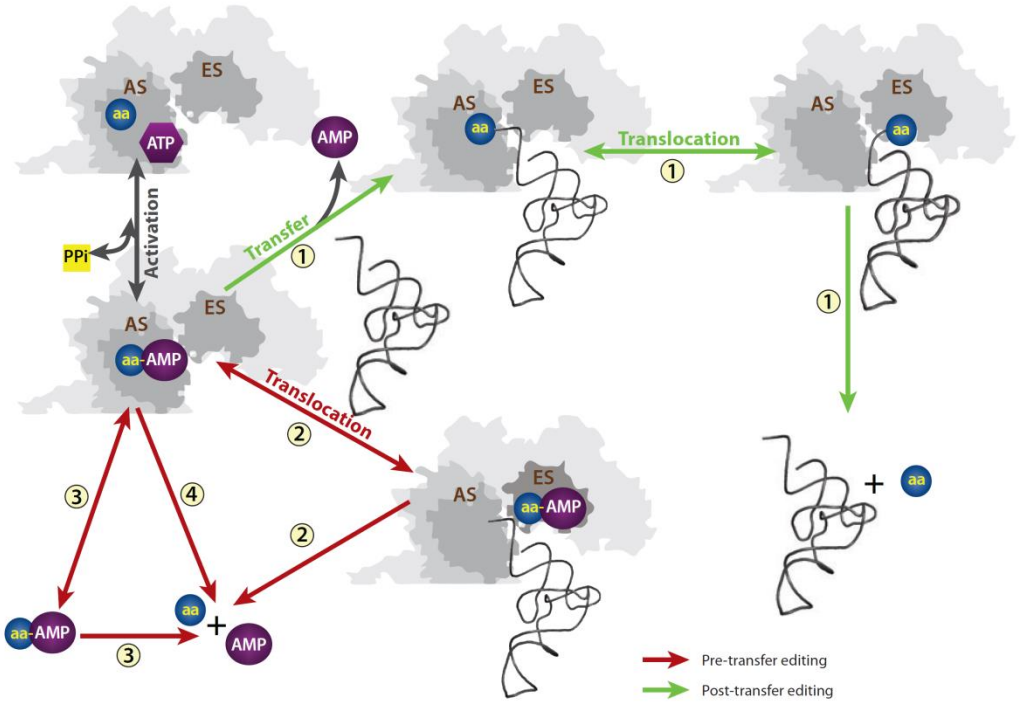


Figure 1-9: aaRS editing pathways. Post-transfer editing. (1) The misactivated amino acid (aa) is linked to tRNA and then translocated from the active site (AS) to the editing site (ES) for hydrolysis. Pre-transfer editing comprises pathways 2 to 4. (2) Translocation of aminoacyl-adenylate (aa-AMP). The misactivated aa-AMP translocates to the ES and is hydrolysed. (3) Selective release. After activation, the non-cognate aa-AMP is released into solution and targeted to spontaneous hydrolysis. (4) Active site hydrolysis. The noncognate aa-AMP is hydrolysed at the AS before release. Adapted from (Ling *et al.*, 2009)

1.3.3. tRNA identity elements and tRNA decoding

The nucleotides or structural elements of a tRNA that are recognized by aaRSs are designated identity determinants. On the other hand, elements that prevent incorrect interactions with noncognate aaRSs are called identity antideterminants (Giege *et al.*, 1998). Determinants and antiderminants are located mainly at the two distal ends of the tRNA, namely the anticodon loop and the acceptor region (Richard Giege and Jacques Lapointe, 2009). More precisely, identity determinants are located in the three first base pairs (1:72, 2:71 and 3:70),

the discriminator base (unpaired 73) and in the anticodon loop, namely anticodon positions 34, 35, 36 and at position 37 (Richard Giege and Jacques Lapointe, 2009) (Figure 1-10).

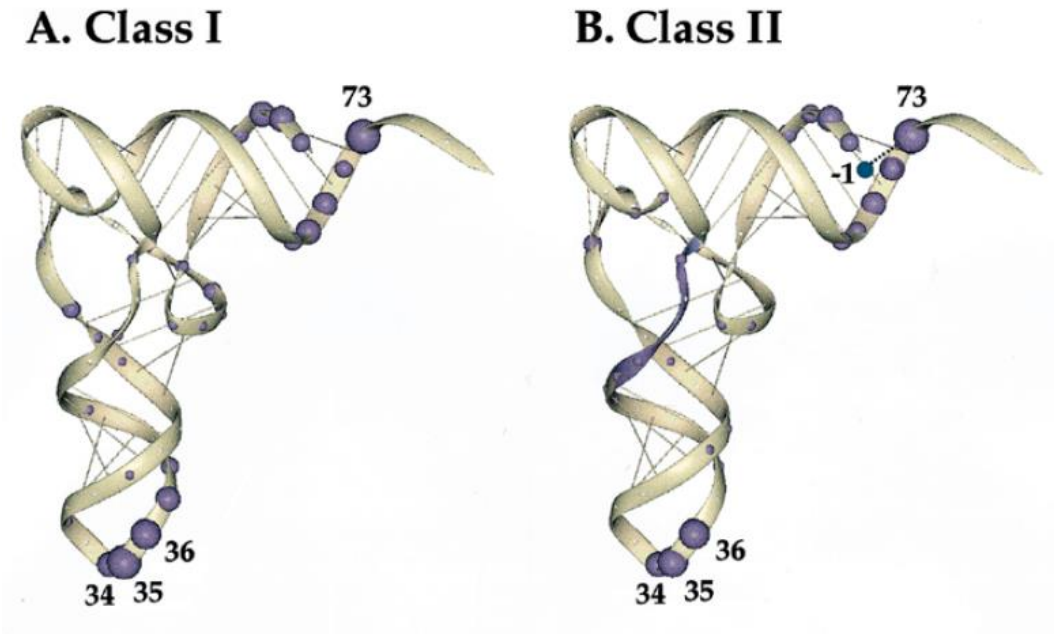


Figure 1-10: Localization of the identity elements on the tRNA structure. Identity elements distribution in tRNA recognized by class I aaRSs (**A**) and class II aaRSs (**B**). The identity elements are mainly located in discriminator base 73, the acceptor stem and the anticodon-loop. The purple spheres size is proportional to the frequency of identity nucleotides at a given position (five increasing sizes of purples color spheres corresponding to 1-fold, 2-3-fold, 4-6-fold, 7-8-fold and 9–10-fold presence of an identity element). Adapted from (Giege *et al.*, 1998).

In general, there are approximately 40 distinct tRNA genes. Therefore, decoding of the 61 amino acid codons require several tRNAs capable of recognizing more than one codon. This decoding mechanism is explained by the Wobble Hypothesis which posits that tRNA nucleotides at positions 35 and 36 base-pair with codons through Watson-Crick base-pairing, however tRNA nucleotide 34 can base-pair with the 3rd codon base through non Watson-Crick base pairing (Crick, 1966). For this reason, the tRNA base 34 is designated the wobble base and modifications in this position alter decoding properties of tRNAs.

The wobble position can be occupied by unmodified nucleotides normally C or G and rarely A or U. Adenine at the wobble position is normally modified into inosine (I) through deamination. Inosine can base pair with U, A and C. On the other hand, uridine generally base-pairs with A or G. But, modified uridine namely

uridine 5-oxyacetcacid (cmo^5U) which is present in alanine, valine and serine tRNAs, is capable of base-pairing with A, G and U and even C. When uridine is modified to 5-methylaminomethyl uridine (mnm^5U) it efficiently base-pairs with A and G, but not with U and C. The 2-thio-modification of uridine (s^2U) present in glutamate, glutamine and lysine tRNAs base-pairs with A or G (Agris *et al.*, 2007;Weixlbaumer and Murphy IV, 2009).

The nucleotide located in position 37 adjacent to the anticodon, the universal purine, is frequently modified. If a codon starts with U or A, nucleotide 37 is a modified adenosine, normally N6-isopentenyladenosine (i6A37) or N6-threonylcarbamoyladenosine (t6A37), respectively (Agris *et al.*, 2007). This nucleotide modification is involved in preventing formation of intra-loop base pairs, maintaining the anticodon loop in the open conformation (Weixlbaumer and Murphy IV, 2009) and enhances base stacking mainly in pyrimidine-rich anticodons (Agris *et al.*, 2007).

1.3.4. Translation

After transcription, mRNAs are extensively processed. Their 5'-end is capped by the addition of m7G cap, the 3'-end id polyadenylated (poly A tail) and introns are removed. The first processing step is the addition of the m_7G cap structure to the 5' end of the nascent mRNA, which occurs after the synthesis of the first 20-30 nucleotides (Gu and Lima, 2005). This process starts by the removal of the ɣ-phosphate from the 5′ end of the nascent transcript by a cap enzyme subunit RNA 5′-triphosphatase and then the other subunit mRNA guanylyltransferase transfers the GMP moiety from GTP to the 5′-diphosphate of the nascent transcript. Next, GMP is methylated at position N7 (Takase *et al.*, 2000). The resulting m7G cap is bound by the two subunits of the cap-binding complex (CBC). The CBC is replaced by translation initiation factor 4E (eIF-4E), an essential step in translation initiation. The m7G cap also plays important roles in mRNA stability and translation (Lackner and Bahler, 2008).

The majority of mRNAs in eukaryotes are interrupted by introns that are removed from the pre-mRNA to produce a functional mRNA, through splicing. For this, consensus sequences marking the exon-intron junctions in the pre-mRNAs are recognized by the spliceosome, a catalytic complex responsible for the enzymatic reactions which removes the introns and ligates the flanking exons (Jurica and Moore, 2003;Lackner and Bahler, 2008). This complex is formed by five small ribonucleoprotein particles (snRNPs), namely U1, U2, U4, U5 and U6, which are made of small nuclear RNA (snRNA) and numerous proteins (Jurica and Moore, 2003). The splicing reaction comprises several types of molecular interactions, namely RNA-protein, RNA-RNA and protein-protein. Additionally, the exons can be alternatively spliced through a process designated as alternative splicing, which leads to the production of multiple protein isoforms and subsequently to greater protein variety (Lackner and Bahler, 2008).

In general, mRNAs also have a poly-A tail at the 3'end. This poly (A) tail is produced by post-transcriptional poly (A) synthesis (Proudfoot and O'Sullivan, 2002). For this, the pre-mRNA is cleaved at an A-rich element and a GU-rich downstream element (DSE), with several polyadenylation complexes involved in the cleavage and in the poly-A tail addition. In yeast, these complexes include the cleavage polyadenylation factor (CPF), which contains the poly (A) polymerase (PAP) and several factors, the cleavage factor IA (CF IA), and cleavage factor IB (CF IB). The nascent poly-A tail is recognized and protected by the poly-A-binding protein (PABP). The poly-A tail is essential for several post-transcriptional regulatory events and its length is crucial for mRNA stability (Lackner and Bahler, 2008). Finally, the mature mRNAs is exported from the nucleus to the cytoplasm for translation.

1.3.4.1. Translation initiation

Initiation is the process that comprises the assembly of the functional 80S ribosome, in which the initiation codon base-pairs with the anticodon loop of the initiator tRNA (Met-$tRNA_i^{Met}$) in the ribosomal P-site. Generally, in eukaryotes the

initiation process comprises a coordinated action of numerous initiation factors, with the exception of some mRNA elements like internal ribosome entry sites (IRESs), which need a subset of initiation factors (Jackson *et al.*, 2010).

This process requires several key steps. The initial step involves the formation of the 43S pre-initiation complex. The eIF2, a methionyl-initiator tRNA (Met-tRNA$_i^{Met}$) and GTP assemble with the aid of eIF2B that promotes the GDP-GTP exchange that reactivates eIF2, forming the ternary complex. Then, the ternary complex binds to the 40S ribosomal subunit with eIF1, eIF3 and eIF1A in mammalian cells, or eIF1, eIF3 and eIF5 in budding yeasts, forming the 43S complex. This complex binds mRNA through interaction of eIF4F, eIF3, eIF4B, and possibly the poly(A)-binding protein (PABP). eIF4F (which contains eIF4E, eIF4G, and eIF4A, except in yeasts) recognizes the m^{7}G cap structure at the 5' end of the mRNA, and PABP associates with the poly(A) tail at the 3' UTR of the mRNA, resulting in a circularized mRNA and regulating the translation initiation process. Subsequently, the 43S complex starts scanning along the 5'-UTR region in the 3' direction, until it reaches the AUG start codon in a favorable context, known as the Kozak sequence (GCC(A/G)CCAUGG). At this point, base-pairing between the anticodon of Met-tRNA$_i^{Met}$ and the start codon forms a stable complex, the 48S complex. Then, the 60S subunit joins to the previous complex forming the 80S ribosome. This is achieved with the aid of eIF5 and the hydrolysis of eIF2-GTP resulting in dissociation of the majority of initiation factors, including eIF2-GDP. A second step of GTP hydrolysis is catalyzed by eIF5B, stimulated by 80S subunit formation, resulting in the eIF5B release (Lackner and Bahler, 2008;Myasnikov *et al.*, 2009;Jackson *et al.*, 2010) (Figure 1-11).

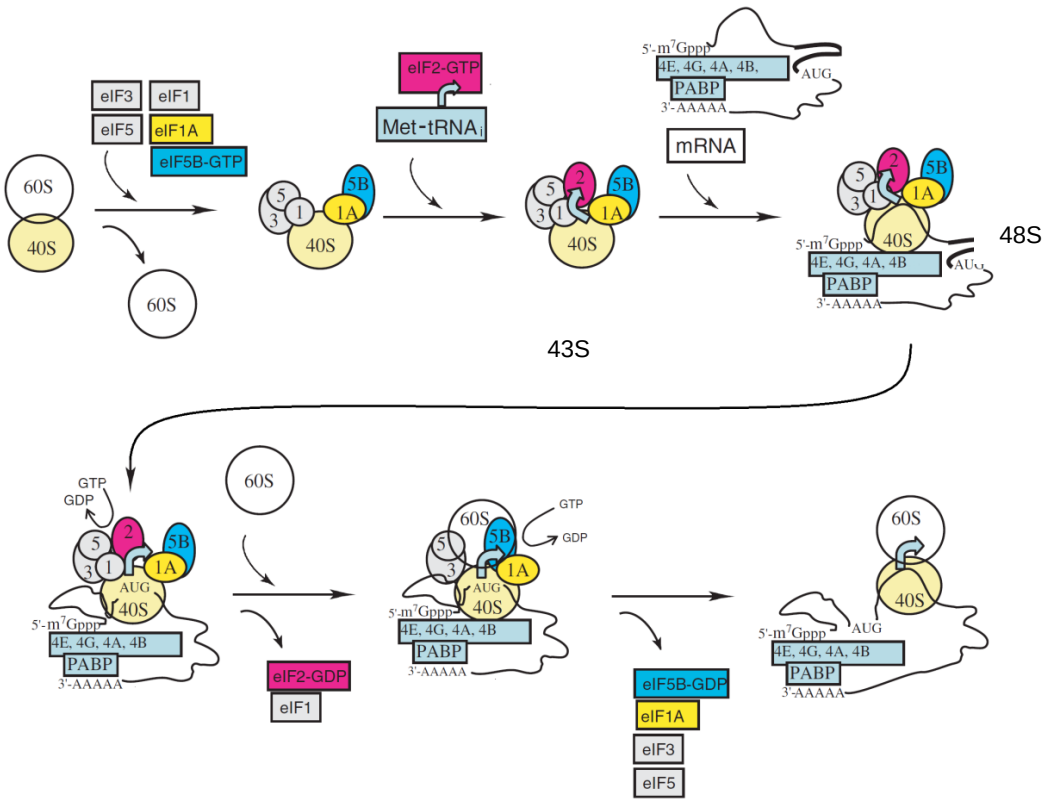


Figure 1-11: Translation initiation. Schematic representation of translation initiation in eukaryotes. After ribosome dissociation, the 43S preinitiation complex is assembled, next this complex attaches to circularized mRNA. After scanning the mRNA until the start codon is reached the assembling of 80S and release of eIF begins, finishing the initiation process. Adapted from (Marintchev and Wagner, 2004)

1.3.4.2. Elongation

The assembling of the 80S ribosome marks the beginning of the elongation phase. The ribosome has three tRNA binding sites, namely, the aminoacyl (A) site, the peptidyl (P) site and the exit (E) site.

During translation elongation, an aminoacyl-tRNA (aa-tRNA) is delivered to the A site, as a component of a ternary complex along with elongation factor 1A (eEF1A) and GTP. When charged tRNAs bind to the A site and a cognate codon-anticodon interaction occurs the GTPase activity on eEF1A is activated, resulting in the release of the aa-tRNA from the ternary complex. Subsequently, eEF1A-GDP is recycled and converted to eEF1A-GTP by eIF1B. The ribosomal peptidyl transferase center catalyzes the formation of the new peptide bond between the

incoming amino acid and the amino acid of the peptidyl-tRNA located in the P-site. This results in a deacylated tRNA (previous peptidyl-tRNA) that moves into the E site, a peptidyl-tRNA which moves to the P site leaving the A site free. Simultaneously, the mRNA is moved three nucleotides placing the next codon at the A site. This process is denominated translocation, is mediated by eEF2 and requires GTP hydrolysis. Since GDP has lower affinity for eEF2 than GTP, GDP is rapidly released and eEF2 binds a new GTP molecule. The tRNA in the E site is released and a new aa-tRNA enters in the A site. This cycle is repeated until a stop codon appears in the A-site, triggering the termination process (Groppo and Richter, 2009;Zaher and Green, 2009) (Figure 1-12).

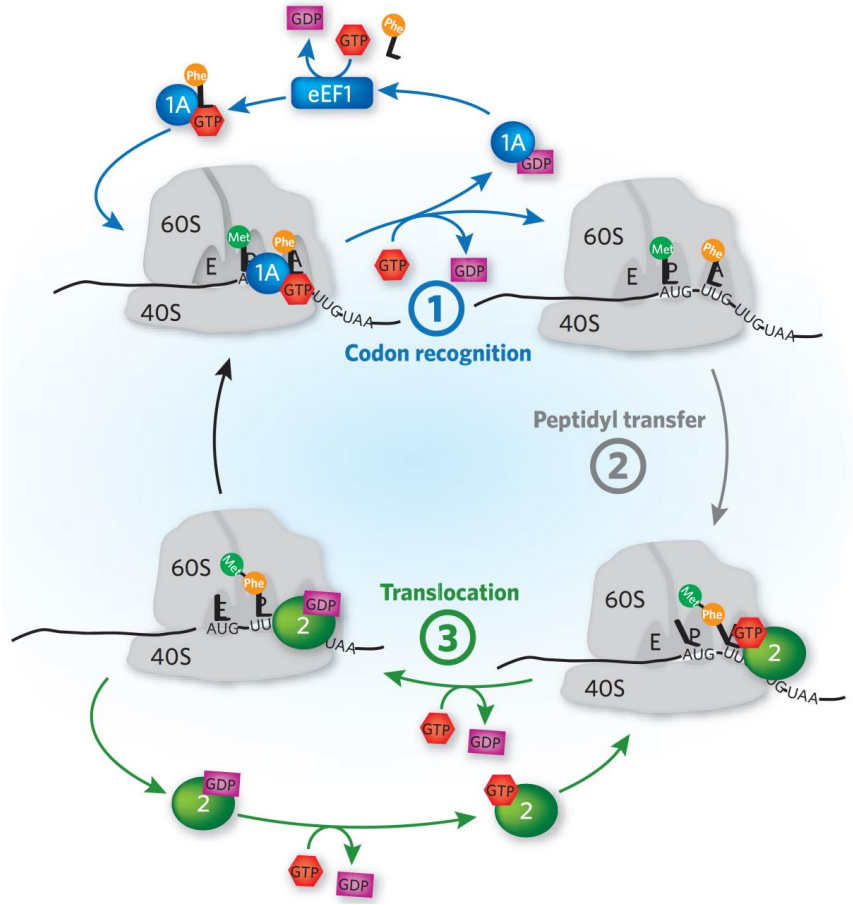


Figure 1-12: Translation elongation. Elongation main steps are highlighted: (1) codon recognition: the aa-tRNA forms a ternary complex with eEF1A and GTP, and binds to the A-site of the ribosome. If the tRNA anticodon matches the mRNA codon, the eEF1A hydrolyses GTP, dissociates from the ribosome and leaves its aa-tRNA behind. (2) peptidyl transfer: The ribosomal peptidyl transferase center catalyses the peptide bond formation, by transferring the nascent peptide chain onto the A-site tRNA, and consequently leaving the P-site tRNA deacylated. (3) translocation: eEF2 moves the peptidyl-tRNA into the P-site and the deacylated tRNA into the E-site, freeing the A-site for another round of elongation. Adapted from (Schneider-Poetsch *et al.*, 2010).

1.3.4.3. Termination

The translation termination process starts when one of the three stop codons, namely UAG, UAA, or UGA, reaches the A-site. The final result of this process is the hydrolysis of the ester bond from the polypeptide with the P site tRNA and the subsequent release of the polypeptide. In eukaryotes, when a stop codon is at the A site, it is recognized by eukaryotic release factor 1 (eRF1). Then, the GTPase eRF3 promotes the release of eRF1 following peptidyl-tRNA hydrolysis (a process that still lacks experimental demonstration) and release of the polypeptide. The final steps in protein synthesis involve the release of the tRNA from the E site and dissociation of the ribosome from mRNA and 40S and 60S subunits dissociation. These processes promote ribosomal subunits recycling, for another round of translation. In eukaryotes, the recycling step is not well understood (Kapp and Lorsch, 2004;Zaher and Green, 2009).

1.3.5. Protein synthesis errors

Proteins synthesis is error prone, with one amino-acid misincorporation occurrence every 1,000 to 10,000 codons translated (Parker, 1989;Ogle and Ramakrishnan, 2005). These errors result from the accumulation of mistakes during protein synthesis, but also include transcription errors, aberrant splicing and premature termination (Figure 1-13) (Drummond and Wilke, 2009). Errors represent a compromise between protein synthesis accuracy and speed (Reynolds *et al.*, 2010).

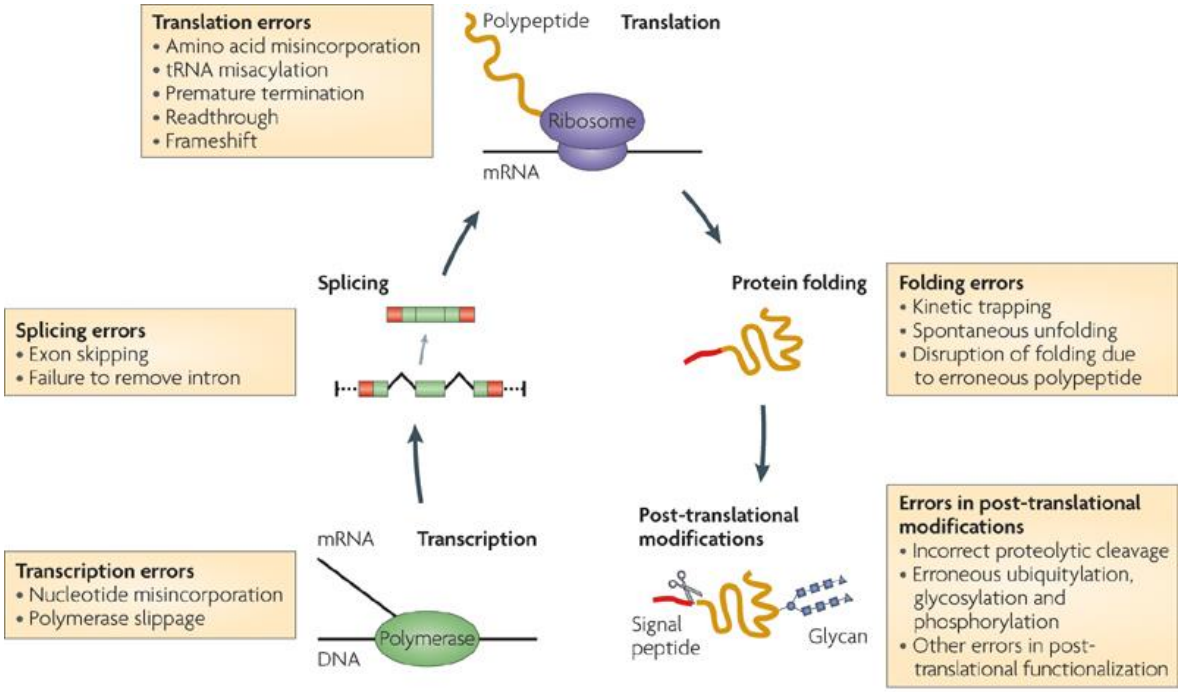


Figure 1-13: Sources of erroneous proteins in eukaryotic cells. Erroneous proteins can have their origin from many stages, namely transcription of DNA into mRNA, splicing processes and mRNA translation. Even after correct translation a faulty post-translational modifications and kinetic misfolding can result in erroneous proteins. Adapted from (Drummond and Wilke, 2009)

Translation errors are divided in two main categories, namely nonsense errors and missense errors (Figure 1-14). Nonsense errors are processivity errors that occur when translation of the coding sequence does not terminate at the first stop codon, originating extended (read-through) polypeptides (Farabaugh and Bjork, 1999;Shah and Gilchrist, 2010). Processivity errors can also include premature termination of translation and frameshifting. Premature termination generally occurs when non-cognate peptidyl-tRNAs dissociate from the P-site because of their weak interaction with the mRNA, originating ribosome drop-off (Menninger, 1977). Another possibility is incorrect recognition of A-site sense codons by the eRF1 (Jorgensen *et al.*, 1993). Translational frameshifting alters the reading frame to the -1 or to the +1 frame by tRNA slippage. Since the reading frame of mRNA is established at the initiation step, it is critical to maintain the 3 bases translocation until a stop codon is reached. The frameshift errors normally occur when a translocation of 2-base (-1 frameshift) or 4-base (+1 frameshift) happens (Parker, 1989). The frequency of these errors is probably bellow 10^{-5} per codon (Farabaugh and Bjork, 1999).

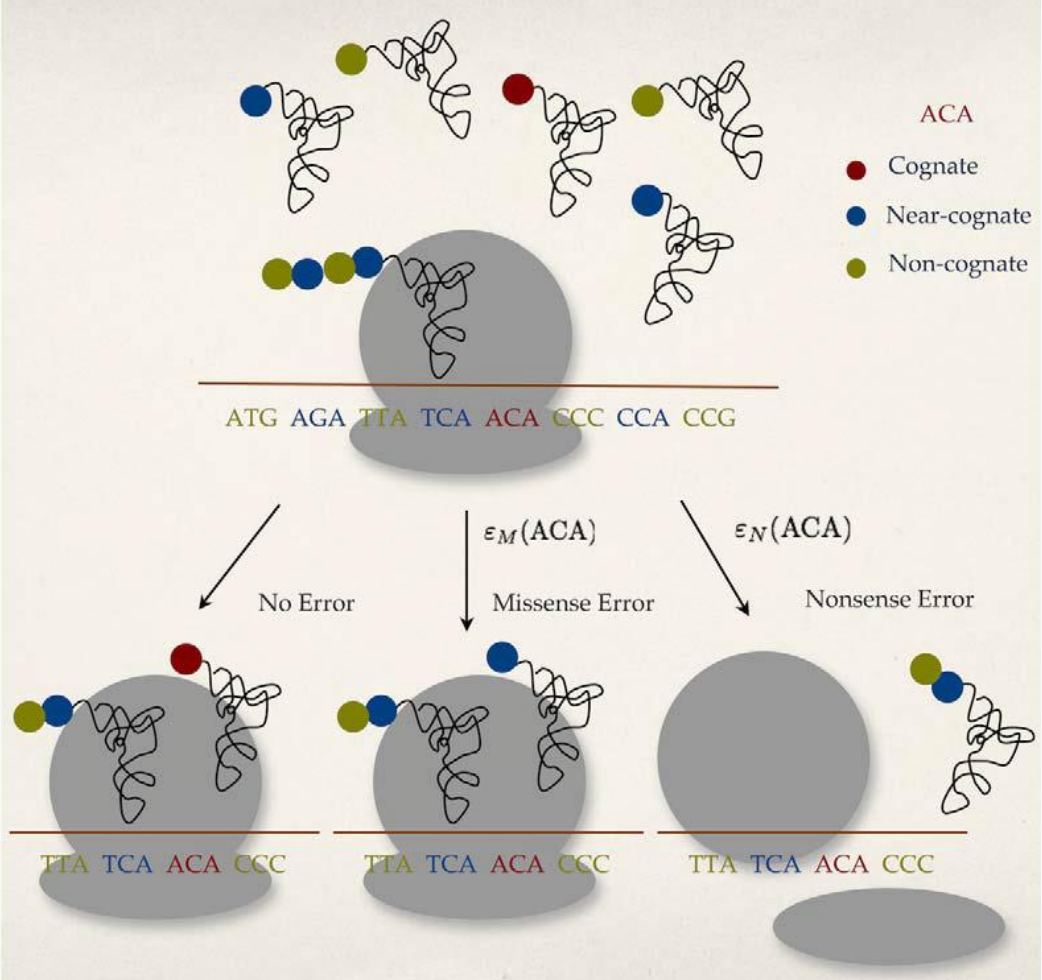


Figure 1-14: Model of translation errors. During translation a ribosome encounters a codon (in this example ACA). One of three possible processes can occur: elongation with a cognate tRNA resulting in correct translation; elongation by a near-cognate tRNA resulting in a missense error; or premature termination of translation due to recognition and a sense codon by a release factor and spontaneous ribosome drop-off or a frameshifting. Adapted from (Shah and Gilchrist, 2010).

On the other hand, missense errors occur when the wrong amino acid is inserted into the growing peptide chain (Shah and Gilchrist, 2010) as a result of a non-cognate interaction between the tRNA anticodon and the mRNA codon, or due to mischarging of tRNA with the wrong amino acid by the amino-acyl tRNA synthetase (Parker, 1989;Stansfield *et al.*, 1998). Protein synthesis errors are generally problematic to organisms and reduce fitness. These fitness costs can arise due to loss of function or gain of toxic function of the erroneous proteins (Drummond and Wilke, 2009). For example, a mouse mutation in the editing domain of the alanyl-tRNA synthetase disrupts the enzyme aminoacylation proofreading activity, generating misacylation and widespread translational errors.

This results in accumulation of misfolded proteins and subsequent death of Purkinje cells in the mouse cerebellum (Lee *et al.*, 2006b).

Several studies also demonstrate that translational errors can be advantageous in particular growth conditions. Engineered *S. cerevisiae* cells mistranslating leucine as serine upregulate heat shock proteins and the general stress response, which results in increased survival rate at 50ºC (Santos *et al.*, 1996) and tolerance to cadmium, arsenate and hydrogen peroxide (Santos *et al.*, 1999). Another example is *C. albicans* where high levels of leucine misincorporation trigger morphogenesis and up-regulate expression of genes related to cell adhesion, hyphal development and increased secretion of proteinases and phospholipases (Miranda *et al.*, 2007), which are potentially advantageous when the pathogen faces host defences.

1.4. Genetic code alterations

The genetic code is a chemical algorithm that translates mRNA codons into amino acids, using the translation machinery, namely mRNA, tRNAs, amino acids, aminoacyl-tRNA synthetases and termination factors. Alterations to the genetic code and its evolution are intrinsically related with the evolution of the components of the translation machinery (Massey *et al.*, 2003;Soll, 2007).

The first discovery of an alteration to the genetic code occurred in 1979, when a report showed that human mitochondria translate the UGA stop codon as tryptophan and that the AUA codon is translated as methionine instead of isoleucine (Barrell *et al.*, 1979). The change in the code was thought to be tolerated in mitochondria because of their relatively small genome (Ohama *et al.*, 2008). However, in 1985 an alteration in the genetic code of *Mycoplasma capricolum* was discovered, namely the use of UGA as a tryptophan rather than a stop codon (Yamao *et al.*, 1985). Currently, variations to the genetic code have been reported in numerous organisms, totalizing 16 codon changes in mitochondrial codes and 10 codon changes in nuclear codes (Knight *et al.*, 2001;Di, 2005).

Although these alterations occur in different organisms, several of them affect the same codons, namely the stop codons UAA, UAG and UGA. In numerous green algae, ciliates (e.g. *Tetrahymena*) and *Diplomonads* the nuclear codons UAA and UAG encode glutamine. The codon UGA encodes: tryptophan in *Mycoplasma* species, *Spiroplasma citri, Bacillus* and some ciliates; cysteine in *Euplotes* and an unidentified amino acid in *Pseudomicrothorax dubius* and *Nyctotherus ovalis* (Soll and RajBhandary, 2006). Some codons are absent and do not code for any amino acid, namely the arginine AGA codon and isoleucine AUA codon in *Micrococcus* species (Kano *et al.*, 1993) and the arginine CGG codon in *Mycoplasma capricolum* (Oba *et al.*, 1991). Other nuclear genetic code alteration is the decoding of CUG as serine in *Candida* species and in many *Ascomycetes* (Santos *et al.*, 1993;Suzuki *et al.*, 1997) (Figure 1-15).

In mitochondria, the genetic code alterations are widespread and a large number of phyla use the stop codon UGA to encode tryptophan, with the exception of green plants. The other stop codon UAA encodes tyrosine in several *Platyhelminths*, while UAG encodes alanine and leucine in fungi and several plants. The codon AAA, which normally encodes lysine, is translated as arginine in *Platyhelminths* and *Echinoderms*. The arginine codons AGA and AGG are termination codons in vertebrates and serine codons in various animal mitochondria. Some organisms encode AGA as glycine and others encode it as serine. The UCA codon is a stop codon rather than serine in the green alga *Scenedesmus obliquus* (Soll and RajBhandary, 2006) (Figure 1-15).

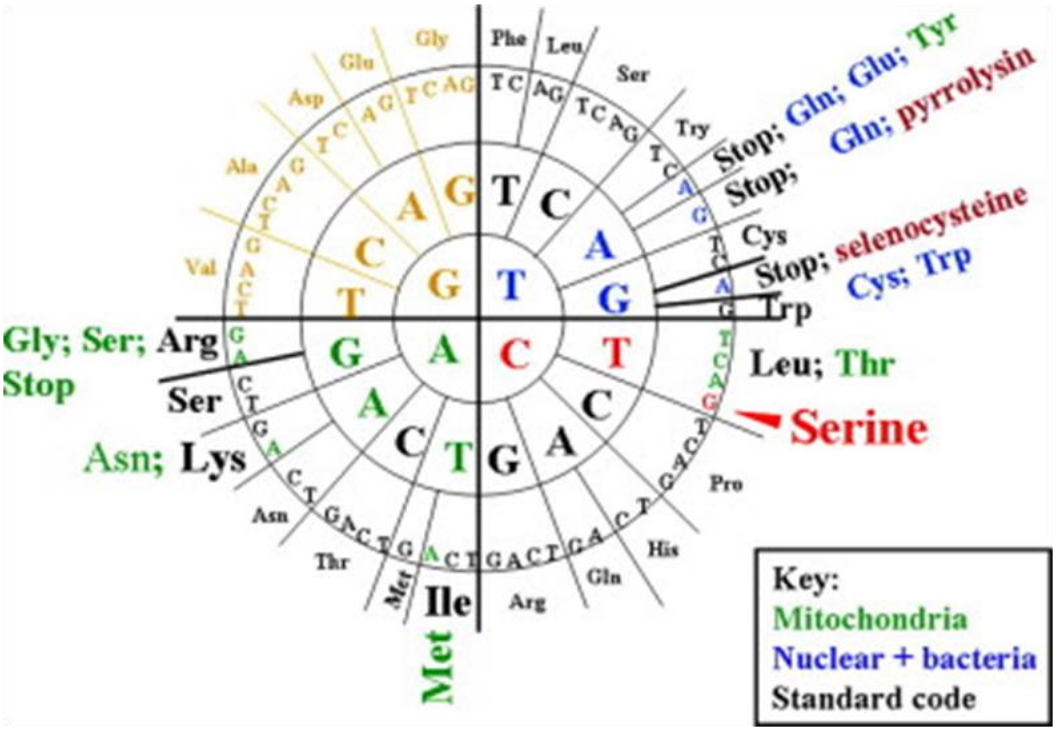


Figure 1-15: Diagram displaying the genetic code alterations discovered to date. Green color indicates genetic code alterations discovered in mitochondria. Blue color indicates genetic code alteration in bacteria and several unicellular eukaryotes. The Stop codons UGA and UAG code selenocysteine and pyrrolysine, respectively. Adapted from (Moura *et al.*, 2010)

These widespread alterations show that the code is neither universal nor frozen and that several codons are more prone to change than others, namely the stop codons UAA, UAG and UGA, the sense codons AUA, AAA, AGA, AGG, CUG and the four-codon boxes CUN and CGN.

1.4.1. Genetic code alteration Theories

1.4.1.1. Codon Capture Theory

The codon capture theory (Figure 1-16) postulates that under genomic G+C or A+T pressure (particularly, a small, e.g., organellar genome), a codon together with the corresponding tRNA disappears. Later, genetic drift may re-introduce the codon if a near-cognate tRNA can decode it. This mechanism of codon reassignment is accomplished without synthesis of aberrant or nonfunctional

proteins, being essentially neutral (Soll and RajBhandary, 2006;Koonin and Novozhilov, 2009).

However, erasing thousands of codons by mutation pressure alone is an improbable evolutionary scenario. Because mutation pressure and genetic drift are weakest evolutionary forces, this must lead to a very slow alteration of genomic composition (Schultz and Yarus, 1996). Furthermore, disappearance of codons driven by G+C pressure is problematic in eukaryotes because their genomes have variable G+C content along chromosomes, providing shelter for rare codons (Santos and Tuite, 1995;Schultz and Yarus, 1996).

1.4.1.2. Ambiguous intermediate

The ambiguous intermediate theory (Figure 1-16) claims that codon reassignment is driven through an intermediate stage where the reassigned codon is ambiguously decoded. This ambiguous decoding may result from competition between two tRNAs, one cognate and one non-cognate (mutant), which may lead to elimination of the gene coding for the cognate tRNA and to reassignment of the codon by the mutant tRNA (Schultz and Yarus, 1996;Koonin and Novozhilov, 2009). A similar mechanism may apply in the reassignment of stop codons to sense codons, however in this case the competition evolves a tRNA and a release factor (RF) (Schultz and Yarus, 1996). The ambiguous intermediate has a negative impact on the survival of the organism due to widespread insertion of wrong amino acids into proteins (Schultz and Yarus, 1996). However, the finding that the CUG codon (normally coding for leucine) is decoded as both leucine (3-5%) or serine (95-97%) in several *Candida* species due to dual charging of a $tRNA_{CAG}$ by the leucyl-tRNA synthetase and seryl-tRNA synthetase, provides strong support for this theory (Suzuki *et al.*, 1997). Also, the fact that mRNA translation has an intrinsic decoding error rate of 10^{-4} (Soll and RajBhandary, 2006) supports this theory.

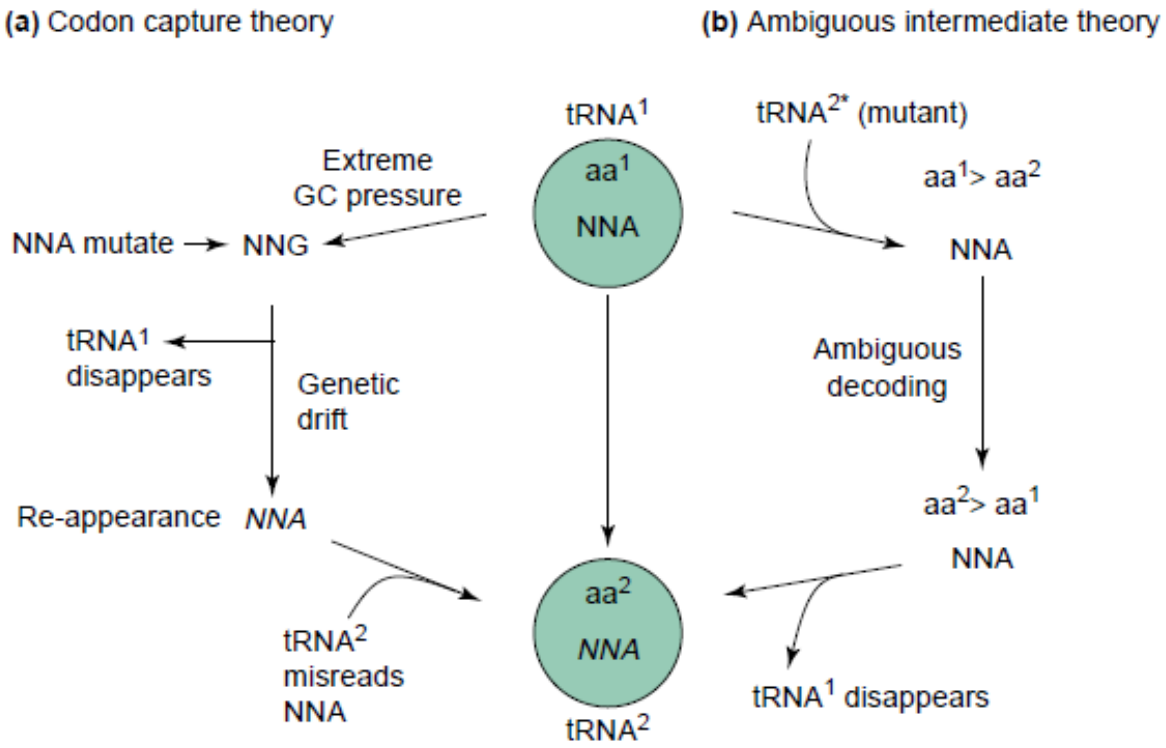


Figure 1-16: Schematic representation of the codon capture and ambiguous intermediate theories. a) Codon capture theory: under pressure by genome G+C bias a codon NNA together with the corresponding tRNA ($tRNA^1$) disappears, the codon re-appears being read by a non-cognate tRNA ($tRNA^2$), changing the codon identity. b) Ambiguous intermediate theory: a tRNA ($tRNA^{2*}$) is mutated and starts to compete with cognate tRNA ($tRNA^1$) for NNS codon decoding, inducing an ambiguous decoding. The disappearance of $tRNA^1$ leads to NNA codon reassignment. Adapted from (Santos *et al.*, 2004)

1.4.1.3. Genome streamlining

The genome streamlining theory postulates that the genetic code changed, at least in animal mitochondria, because there is selective pressure to minimize the mitochondrial translation apparatus, due to genome size reduction. This results in the reassignment of specific codons, being the stop codons one example (Andersson and Kurland, 1998;Soll and RajBhandary, 2006;Koonin and Novozhilov, 2009). This theory is supported by the observation that animal mitochondria diverged significantly from their ancestors, leading to a specialized tRNA set (Andersson and Kurland, 1998). To translate the universal genetic code in the 20 proteinaceous amino acids, the minimal set of different tRNAs needed is 24. However, in some animal mitochondria, the number of tRNA genes is 22. This

number corresponds to several rearrangements. For example, the isoleucine-tRNA for the AUA codon and the arginine-tRNA for the AGA and AGG codons are missing. Moreover, the methionine-tRNA translates AUG and AUA codons. And, AGA and AGG are termination codons or translated by a serine–tRNA that reads AGN codons (Andersson and Kurland, 1998).

1.4.2. Natural genetic code expansions

In recent years, two new amino acids were added to the canonical 20 amino acids of the genetic code. These amino acids are assigned to termination codons. Selenocysteine, also known as the 21st amino acid, is found in various species from the 3 life kingdoms of life and is encoded by the UGA codon (Bock *et al.*, 1991). The 22st amino acid is pyrrolysine and is encoded by the UAG codon and in *Methanosarcinaceae* (Srinivasan *et al.*, 2002;Soll and RajBhandary, 2006).

1.4.2.1. Selenocysteine

Selenocysteine and selenoproteins are found in all domains of life, although some organisms including some fungi and higher plants do not use it. Mammals have at least 20 selenoproteins, and several archaea species, namely *Methanocaldococcus jannaschii*, *Methanococcus maripaludis* and *Methanopyrus kandleri* also encode it (Ambrogelly *et al.*, 2007). Selenocysteine differs from cysteine by selenium replacement at the sulphur atom (Allmang and Krol, 2006). The most well-known selenoproteins are oxidoreductases that have selenocysteine in the active site. Chemically, selenium is similar to sulphur but has stronger antioxidant capacity (Allmang and Krol, 2006;Ambrogelly *et al.*, 2007).

In Bacteria, biosynthesis of selenocysteine (Figure 1-17) starts with acylation of the tRNASec(SelC) with serine by the SerRS, although the tRNASec atypical structures dramatically decrease SerRS recognition efficiency (1% when compared with the tRNASer). Then, Ser-tRNASec reacts with selenocysteine synthase (SelA), a pyridoxal phosphate (PLP)-dependent enzyme with a

conserved lysine residue. This enzyme is responsible for the formation of an enzyme-bound dehydroalanyl intermediate with elimination of a water molecule, which results in SelA-bound dehydroalanyl-tRNASec. After that, Selenophosphate synthetase (SelD) produced from selenide and ATP, and a monoselenophosphate (a selenium donor) react with SelA-bound dehydroalanyl-tRNASec to form Sec-tRNA (Ambrogelly *et al.*, 2007).

In bacteria, recoding of the UGA codon to incorporate selenocysteine requires the presence of a mRNA selenocysteine insertion sequence (SECIS) stem-loop structure and specialized translation elongation factor (SelB). The SECIS element is located downstream of the UGA codon in the selenoprotein coding region. The SelB N-terminal binds to Sec-tRNA^{Sec} with high specificity and the C-terminal binds to SECIS element allowing for delivery of Sec-tRNA^{Sec} to ribosomal A site by overcoming the release-factor binding to the UGA codon (Ambrogelly *et al.*, 2007).

In eukaryotes and archaea, biosynthesis of selenocysteine (Figure 1-17) is apparently a tRNA-dependent conversion of serine, a selenocysteinyl-tRNA synthetase has never been discovered. In fact, serylation of tRNASec by SerRS has been demonstrated *in vitro* in archaea and eukaryotes and is assumed to occur *in vivo* (Ambrogelly *et al.*, 2007). Subsequently, Ser-tRNA^{Sec} is phosphorylated by *O*-phosphoseryl-tRNA kinase (PSTK) forming *O*-phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}) (Yuan *et al.*, 2010). Then, Sep-tRNA:Sec-tRNA synthase (SepSecS) reacts with selenophosphate and Sep-tRNA^{Sec} and form the final product Sec-tRNA^{Sec} (Yuan *et al.*, 2010). The recoding of UGA codon in archaea is not well understood, although several models have been proposed (Allmang and Krol, 2006).

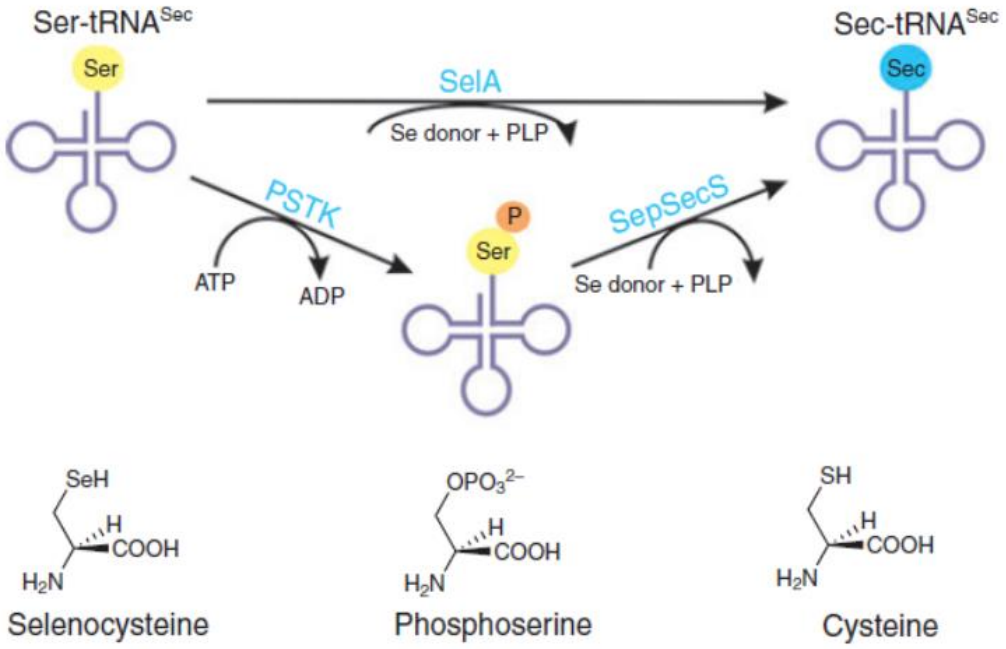


Figure 1-17: Scheme showing the formation of Sec-tRNA^{Sec}. Top) the SelA pathway found in bacteria. Ser-tRNASec reacts with selenocysteine synthase (SelA) and a monoselenophosphate (a selenium donor) to form Sec-tRNA^{Sec}. The PSTK/SepSecS pathway is found in archaea and eukaryotes. Ser-tRNA^{Sec} is phosphorylated by *O*-phosphoseryl-tRNA kinase (PSTK) forming *O*-phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}). Then, Sep-tRNA:Sec-tRNA synthase (SepSecS) reacts with selenophosphate and Sep-tRNA^{Sec} and form the final product Sec-tRNA^{Sec}. Selenocysteine, *O*-phosphoserine and cysteine, chemical structures are displayed on the bottom. Adapted from (Ambrogelly *et al.*, 2007)

1.4.2.2. Pyrrolysine

The 22nd amino acid (pyrrolysine) is found in *Methanosarcinaceae* and in the bacterium *Desulfitobacterium hafniense* (Ambrogelly *et al.*, 2007). This amino acid is required for methane formation from methylamines (Gaston *et al.*, 2011b). Pyrrolysine chemical structure is characterized by a methylated pyrroline carboxylate in amide linkage to the ε-amino group of L-lysine2 (Figure 1-18) (Yuan *et al.*, 2010;Gaston *et al.*, 2011b).

Gaston and colleagues showed that an *E. coli* transformed with *pylTSBCD* genes (from methanogenic archaea) is able to incorporate endogenously biosynthesized pyrrolysine into protein (Gaston *et al.*, 2011b). These authors were

able to demonstrate a *pylBCD*-dependent pathway in which pyrrolysine arises from two lysines (Gaston *et al.*, 2011b). The recoding of UAG as pyrrolysine requires *pylT* that produces $tRNA^{Pyl}$ ($tRNA_{CUA}$) and pyrrolysyl-tRNA synthetase (PylRS) encoded by *pylS.* The PylRS is considered the 21st AARS, since The $tRNA^{Pyl}$ ($tRNA^{Pyl}_{CUA}$) is aminoacylated directly by the pyrrolysyl-tRNA synthetase (PylRS) (Gaston *et al.*, 2011a;Gaston *et al.*, 2011b).

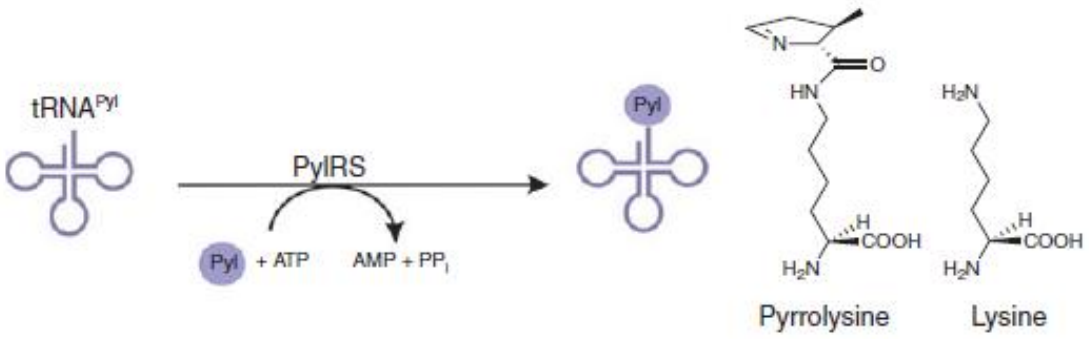


Figure 1-18: Scheme showing the formation of Pyl-$tRNA^{Pyl}$. The $tRNA^{Pyl}$ ($tRNA^{Pyl}_{CUA}$) is aminoacylated directly by the pyrrolysyl-tRNA synthetase (PylRS). Pyrrolysine and lysine chemical structures are display on the side. Adapted from (Ambrogelly *et al.*, 2007).

1.5. *Candida albicans*

Candida albicans belongs to the *Saccharomycetaceae* family of ascomycete fungi. It is an opportunistic pathogen with a remarkable capacity to colonize numerous sites in the human body. In the commensal state, *C. albicans* is part of healthy individuals microflora causing no disease, occupying the oral cavity, gastrointestinal tract, genitourinary tract, anus and skin. When the host physiology changes and the immune defense becomes compromised, *C. albicans* may cause infection at the already occupied sites and in extreme scenarios it may originate candidemia, which is a blood stream infection and subsequently systemic or disseminated candidiasis with the occurrence of *Candida* in numerous organs, such as the liver, eyes, kidneys, heart, brain and joints (Soll, 2002;Kim and Sudbery, 2011).

Another remarkable characteristic of *C. albicans* is the morphological diversity of its cells, varying from unicellular yeast like cells to pseudohyphae and true hyphae (Sudbery *et al.*, 2004). Unicellular yeast cells grow through budding while true hyphae are like a chain of tubes without constrictions at the septal junctions, pseudohypha cells form daughter buds that elongate and after the formation of septa they fail to separate, forming a filament of cells, although bud elongation can be so extreme that resembles hyphae (Figure 1-19) (Sudbery *et al.*, 2004).This morphological variability is important for *C. albicans* pathogenicity since both pseudohyphae and hyphae can invade the agar substratum when grown in the laboratory, which might be important for tissue penetration and organ colonization. The yeast form is more prone for bloodstream infections (Sudbery *et al.*, 2004;Kim and Sudbery, 2011).

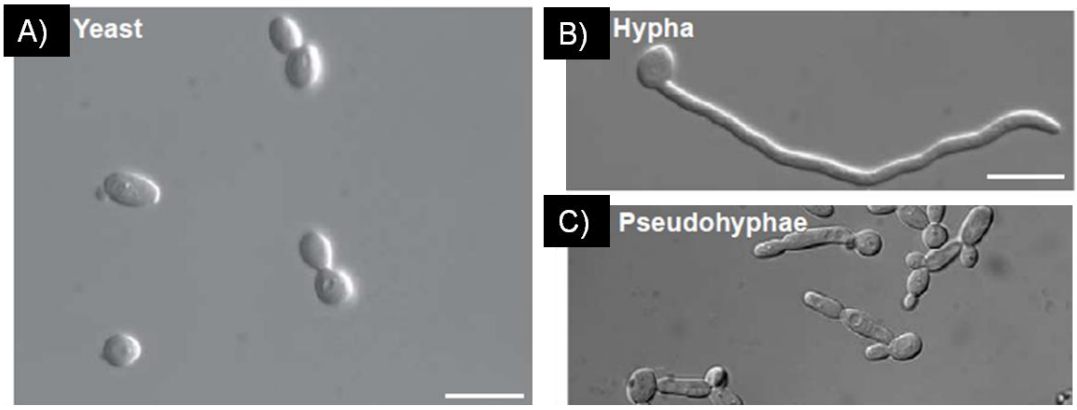


Figure 1-19: *C. albicans* morphologies. (A) Budding yeast cells, (B) Hyphae (C) Pseudohyphae. Adapted from (Sudbery *et al.*, 2004).

The transition between morphological forms is promoted by a variety of environmental conditions. The yeast form is favored at acidic pH (pH 4) and 30ºC. Hyphal growth is promoted at 37ºC in presence of serum, neutral pH, high CO_2 levels growth in embedded conditions, and N-acetylglucosamine. Finally, pseudohyphal growth occurs at 35ºC or pH 5.5, which represents the shifting point between hyphal and yeast growth (Kim and Sudbery, 2011). The transition from yeast to hyphal forms is controlled by a quorum-sensing mechanism that consists in the production of the alcohol farnesol by yeast cells (Hornby *et al.*, 2001).

1.5.1. The *Candida albicans* genetic code

C. albicans is related to the eukaryotic model *Saccharomyces cerevisiae* but has a distinctive characteristic namely the ambiguous translation of its CUG codon (Suzuki *et al.*, 1997). This codon has been reassigned from Leu to Ser, but it is still generally translated ambiguously as serine (95-97%) and leucine (2.96% to 4.95%) (Gomes *et al.*, 2007). This ambiguous translation is mediated by a unique $tRNA_{CAG}^{Ser}$ (Figure 1-20), which is recognized by both the SerRS and the LeuRS. In other words, in *C. albicans* cells there are two forms of the $tRNA_{CAG}^{Ser}$, the Ser-$tRNA_{CAG}^{Ser}$ (charged by the SerRS) and the Leu-$tRNA_{CAG}^{Ser}$ (charged by the LeuRS). Since Leu-$tRNA_{CAG}^{Ser}$ is not edited by the LeuRS nor discriminated by the translation elongation factor 1 (eEF1A) (Santos *et al.*, 1997;Santos *et al.*, 2011), this aminoacylated tRNA competes for CUG codons during translation and incorporates Leu at low level at Ser-CUG positions (Gomes *et al.*, 2007).

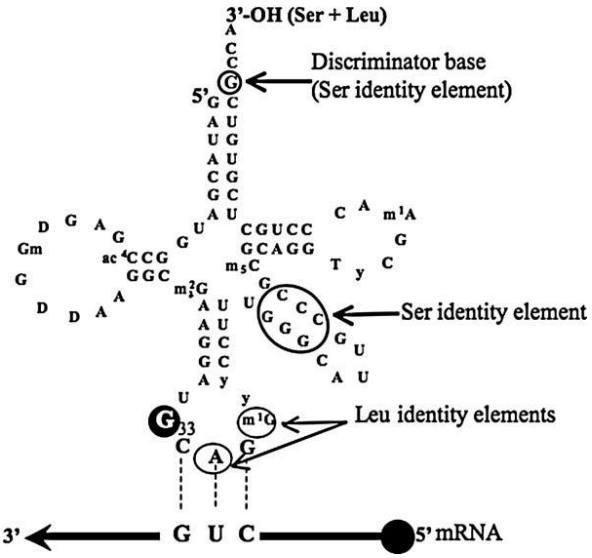


Figure 1-20: Secondary structure of the *C. albicans* $tRNA_{CAG}^{Ser}$. The $tRNA_{CAG}^{Ser}$ is an hybrid tRNA, which is recognized by both the LeuRS and the SerRS. Adapted from (Santos *et al.*, 2011)

The $tRNA_{CAG}^{Ser}$ identity elements recognized by the SerRS are the discriminator base (G_{73}) and three G-C pairs in the variable arm (Suzuki *et al.*, 1997). The LeuRS recognizes the A_{35} and m^1G_{37} in the anticodon loop of the $tRNA_{CAG}^{Ser}$ (Soma *et al.*, 1996). Another interesting feature of the $tRNA_{CAG}^{Ser}$ is the presence of a guanosine at position 33 (G_{33}) instead of the critical uridine (U_{33}),

which is required for the functional conformation of the anticodon U-turn and for the correct stacking of the anticodon bases (Ashraf *et al.*, 1999). This G_{33} mutation may have played an important role in CUG reassignment since it lowered decoding and leucylation efficiencies (Santos *et al.*, 1996;Suzuki *et al.*, 1997). *C. albicans* shares this unique CUG translation system with other fungi of the so called CTG clade (Figure 1-21) (Butler *et al.*, 2009).

Another important characteristic of *C. albicans* is its diploid status, its genome has 33.5% of G+C is organized in eight chromosomes, 13.3-13.4 Mb and encodes 6202 genes. Comparative genomics studies showed that CTG clade species expanded the number of genes encoding transcription factors, lipases, oligopeptide transporters, cell wall mannoproteins, and ferric reductases (Butler *et al.*, 2009).

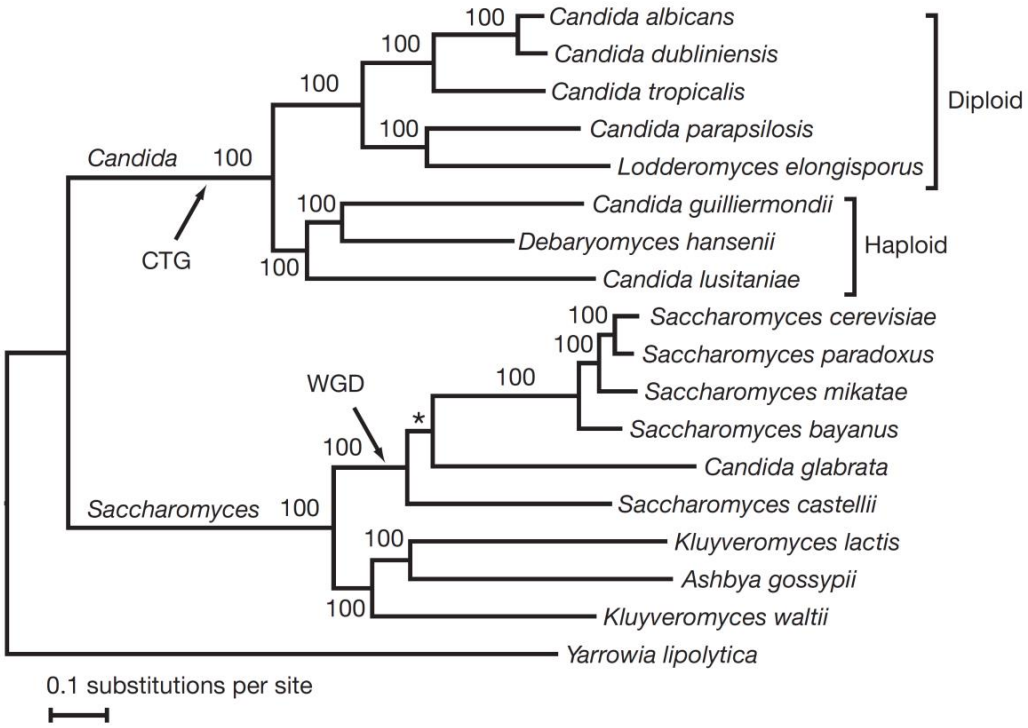


Figure 1-21: Phylogeny of *Candida* and *Saccharomyces* clade species. Adapted from (Butler *et al.*, 2009).

1.6. Aims of this study

C. albicans mistranslates the CUG codon at levels that are lethal to other organisms. In this thesis we aimed at elucidating the *C. albicans* ambiguous translation system. For this, we have developed a GFP reporter system to monitor leucine incorporation at CUG codons. We have also engineered *C. albicans* strains to ambiguously translate CTA , CTC, CTT (Leucine), ATC (Isoleucine), GCC (Alanine), GGA (Glycine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) as serine, to determine whether the genetic code flexibility observed in the CUG codon also exists in other codons.

The specific biological questions addressed in this PhD thesis were the following:

1. Is CUG ambiguity fixed or can it fluctuate in response to environmental variation?
2. What is the nature of the molecular mechanism that allows for variable CUG ambiguity?
3. Is *C. albicans* able to tolerate ambiguity at other codons? Do *C. albicans* cells respond to ambiguity at CUG and non-CUG codons in similar ways?

Our results show that CUG ambiguity is dynamic, that *C. albicans* tolerates very high levels of leucine incorporation at CUG codons and that *C. albicans* and *S. cerevisiae* respond in a totally different manner to mRNA mistranslation. The Data provide strong evidence for the existence of a unique stress response mechanism in *C. albicans*.

1.7. References

Abelson,J., Trotta,C.R., and Li,H. (1998). tRNA splicing. J. Biol. Chem. *273*, 12685-12688.

Agris,P.F. (2004). Decoding the genome: a modified view. Nucleic Acids Res. *32*, 223-238.

Agris,P.F., Vendeix,F.A., and Graham,W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. J. Mol. Biol. *366*, 1-13.

Allmang,C. and Krol,A. (2006). Selenoprotein synthesis: UGA does not end the story. Biochimie *88*, 1561-1571.

Ambrogelly,A., Palioura,S., and Soll,D. (2007). Natural expansion of the genetic code. Nat. Chem. Biol. *3*, 29-35.

Andersson,S.G. and Kurland,C.G. (1998). Reductive evolution of resident genomes. Trends Microbiol. *6*, 263-268.

Angellotti,M.C., Bhuiyan,S.B., Chen,G., and Wan,X.F. (2007). CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Res. *35*, W132-W136.

Arnez,J.G. and Moras,D. (1997). Structural and functional considerations of the aminoacylation reaction. Trends Biochem. Sci. *22*, 211-216.

Ashraf,S.S., Ansari,G., Guenther,R., Sochacka,E., Malkiewicz,A., and Agris,P.F. (1999). The uridine in "U-turn": contributions to tRNA-ribosomal binding. RNA. *5*, 503-511.

Barrell,B.G., Bankier,A.T., and Drouin,J. (1979). A different genetic code in human mitochondria. Nature *282*, 189-194.

Berleant,D., White,M., Pierce,E., Tudoreanu,E., Boeszoermenyi,A., Shtridelman,Y., and Macosko,J.C. (2009). The genetic code--more than just a table. Cell Biochem. Biophys. *55*, 107-116.

Bock,A., Forchhammer,K., Heider,J., Leinfelder,W., Sawers,G., Veprek,B., and Zinoni,F. (1991). Selenocysteine: the 21st amino acid. Mol. Microbiol. *5*, 515-520.

Butler,G. *et al.* (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature *459*, 657-662.

Cavarelli,J. and Moras,D. (1993). Recognition of tRNAs by aminoacyl-tRNA synthetases. FASEB J. *7*, 79-86.

Chernyakov,I., Whipple,J.M., Kotelawala,L., Grayhack,E.J., and Phizicky,E.M. (2008). Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. Genes Dev. *22*, 1369-1380.

Crick,F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. J. Mol. Biol. *19*, 548-555.

Crick,F.H. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367-379.

Czerwoniec,A., Dunin-Horkawicz,S., Purta,E., Kaminska,K.H., Kasprzak,J.M., Bujnicki,J.M., Grosjean,H., and Rother,K. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. Nucleic Acids Res. *37*, D118-D121.

Di,G.M. (2005). The origin of the genetic code: theories and their relationships, a review. Biosystems *80*, 175-184.

Drummond,D.A. and Wilke,C.O. (2009). The evolutionary consequences of erroneous protein synthesis. Nat. Rev. Genet. *10*, 715-724.

Farabaugh,P.J. and Bjork,G.R. (1999). How translational accuracy influences reading frame maintenance. EMBO J. *18*, 1427-1434.

Francklyn,C.S. (2008). DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression. Biochemistry *47*, 11695-11703.

Gaston,M.A., Jiang,R., and Krzycki,J.A. (2011a). Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr. Opin. Microbiol. *14*, 342-349.

Gaston,M.A., Zhang,L., Green-Church,K.B., and Krzycki,J.A. (2011b). The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature *471*, 647-650.

Giege,R. (2008). Toward a more complete view of tRNA biology. Nat. Struct. Mol. Biol. *15*, 1007-1014.

Giege,R., Sissler,M., and Florentz,C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. *26*, 5017-5035.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206.

Groppo,R. and Richter,J.D. (2009). Translational control from head to tail. Curr. Opin. Cell Biol. *21*, 444-451.

Gu,M. and Lima,C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Curr. Opin. Struct. Biol. *15*, 99-106.

Gu,W., Zhou,T., Ma,J., Sun,X., and Lu,Z. (2004). The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. Biosystems *73*, 89-97.

Guo,M. and Schimmel,P. (2012). Structural analyses clarify the complex control of mistranslation by tRNA synthetases. Curr. Opin. Struct. Biol. *22*, 119-126.

Haig,D. and Hurst,L.D. (1991). A quantitative measure of error minimization in the genetic code. J. Mol. Evol. *33*, 412-417.

Hieronim Jakubowski (2005). Accuracy of Aminoacyl-TRNA Synthetases: Proofreading of Amino Acids. In: The Aminoacyl-tRNA Synthetases, ed. Michael Ibba, Christopher Francklyn, and Stephen CusackGEORGETOWN, TEXAS: LANDES BIOSCIENCE / EUREKAH.COM, 384-397.

Hopper,A.K., Pai,D.A., and Engelke,D.R. (2010). Cellular dynamics of tRNAs and their genes. FEBS Lett. *584*, 310-317.

Hornby,J.M., Jensen,E.C., Lisec,A.D., Tasto,J.J., Jahnke,B., Shoemaker,R., Dussault,P., and Nickerson,K.W. (2001). Quorum sensing in the dimorphic fungus *Candida albicans* is mediated by farnesol. Appl. Environ. Microbiol. *67*, 2982-2992.

Houseley,J. and Tollervey,D. (2009). The many pathways of RNA degradation. Cell *136*, 763-776.

Jackson,R.J., Hellen,C.U., and Pestova,T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113-127.

Jakubowski,H. (2011). Quality control in tRNA charging -- editing of homocysteine. Acta Biochim. Pol. *58*, 149-163.

Jorgensen,F., Adamski,F.M., Tate,W.P., and Kurland,C.G. (1993). Release factor-dependent false stops are infrequent in *Escherichia coli*. J. Mol. Biol. *230*, 41-50.

Jurica,M.S. and Moore,M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell *12*, 5-14.

Kano,A., Ohama,T., Abe,R., and Osawa,S. (1993). Unassigned or nonsense codons in *Micrococcus luteus*. J. Mol. Biol. *230*, 51-56.

Kapp,L.D. and Lorsch,J.R. (2004). The molecular mechanics of eukaryotic translation. Annu. Rev. Biochem. *73*, 657-704.

Kim,J. and Sudbery,P. (2011). *Candida albicans*, a major human fungal pathogen. J. Microbiol. *49*, 171-177.

Knight,R.D., Freeland,S.J., and Landweber,L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet *2*, 49-58.

Koonin,E.V. and Novozhilov,A.S. (2009). Origin and evolution of the genetic code: the universal enigma. IUBMB. Life *61*, 99-111.

Lackner,D.H. and Bahler,J. (2008). Translational control of gene expression from transcripts to transcriptomes. Int. Rev. Cell Mol. Biol. *271*, 199-251.

Lee,J.W., Beebe,K., Nangle,L.A., Jang,J., Longo-Guess,C.M., Cook,S.A., Davisson,M.T., Sundberg,J.P., Schimmel,P., and Ackerman,S.L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature *443*, 50-55.

Ling,J., Reynolds,N., and Ibba,M. (2009). Aminoacyl-tRNA synthesis and translational quality control. Annu. Rev. Microbiol. *63*, 61-78.

Marck,C., Kachouri-Lafond,R., Lafontaine,I., Westhof,E., Dujon,B., and Grosjean,H. (2006). The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. Nucleic Acids Res. *34*, 1816-1835.

Marintchev,A. and Wagner,G. (2004). Translation initiation: structures, mechanisms and evolution. Q. Rev. Biophys. *37*, 197-284.

Massey,S.E., Moura,G., Beltrao,P., Almeida,R., Garey,J.R., Tuite,M.F., and Santos,M.A. (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida spp*. Genome Res. *13*, 544-557.

Menninger,J.R. (1977). Ribosome editing and the error catastrophe hypothesis of cellular aging. Mech. Ageing Dev. *6*, 131-142.

Mitchell, M. and Li, H. (2009). Molecular Basis of TRNA Processing Reactions. In: DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution, ed. Henri GrosjeanAustin, Texas, USA: Landes Bioscience, 328-346.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A Genetic Code Alteration Is a Phenotype Diversity Generator in the Human Pathogen *Candida albicans*. PLoS. ONE. *2*, e996.

Moura,G.R., Paredes,J.A., and Santos,M.A. (2010). Development of the genetic code: insights from a fungal codon reassignment. FEBS Lett. *584*, 334-341.

Myasnikov,A.G., Simonetti,A., Marzi,S., and Klaholz,B.P. (2009). Structure-function insights into prokaryotic and eukaryotic translation initiation. Curr. Opin. Struct. Biol. *19*, 300-309.

Nakanishi,K. and Nureki,O. (2005). Recent progress of structural biology of tRNA processing and modification. Mol. Cells *19*, 157-166.

Nirenberg,M. (2004). Historical review: Deciphering the genetic code--a personal account. Trends Biochem. Sci. *29*, 46-54.

Nirenberg,M. and Leder,P. (1964). RNA codewords and protein synthesis. the effect of trinucleotides upon the binding of sRNA to ribosomes. Science *145*, 1399-1407.

Nirenberg,M.W. and Matthaei,J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. U. S. A *47*, 1588-1602.

Oba,T., Andachi,Y., Muto,A., and Osawa,S. (1991). CGG: an unassigned or nonsense codon in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. U. S. A *88*, 921-925.

Ogle,J.M. and Ramakrishnan,V. (2005). Structural insights into translational fidelity. Annu. Rev. Biochem. *74*, 129-177.

Ohama,T., Inagaki,Y., Bessho,Y., and Osawa,S. (2008). Evolving genetic code. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *84*, 58-74.

Parker,J. (1989). Errors and alternatives in reading the universal genetic code. Microbiol. Rev. *53*, 273-298.

Phizicky,E.M. and Hopper,A.K. (2010). tRNA biology charges to the front. Genes Dev. *24*, 1832-1860.

Proudfoot,N. and O'Sullivan,J. (2002). Polyadenylation: a tail of two complexes. Curr. Biol. *12*, R855-R857.

Reynolds,N.M., Lazazzera,B.A., and Ibba,M. (2010). Cellular mechanisms that control mistranslation. Nat. Rev. Microbiol. *8*, 849-856.

Ribas de Pouplana,L. and Schimmel,P. (2001a). Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem. Sci. *26*, 591-596.

Ribas de Pouplana,L. and Schimmel,P. (2001b). Two Classes of tRNA Synthetases Suggested by Sterically Compatible Dockings on tRNA Acceptor Stem. Cell *104*, 191-193.

Richard Giege and Jacques Lapointe (2009). Transfer RNA Aminoacylation and Modified Nucleosides. In: DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution, ed. Henri GrosjeanAustin, Texas, USA: Landes Bioscience, 475-492.

Ronneberg,T.A., Landweber,L.F., and Freeland,S.J. (2000). Testing a biosynthetic theory of the genetic code: Fact or artifact? Proc. Natl. Acad. Sci. U. S. A. *97*, 13690-13695.

Santos,M.A., Cheesman,C., Costa,V., Moradas-Ferreira,P., and Tuite,M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida spp*. Mol. Microbiol. *31*, 937-947.

Santos,M.A., Gomes,A.C., Santos,M.C., Carreto,L.C., and Moura,G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607-611.

Santos,M.A., Keith,G., and Tuite,M.F. (1993). Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. EMBO J. *12*, 607-616.

Santos,M.A., Moura,G., Massey,S.E., and Tuite,M.F. (2004). Driving change: the evolution of alternative genetic codes. Trends Genet. *20*, 95-102.

Santos,M.A., Perreau,V.M., and Tuite,M.F. (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. EMBO J. *15*, 5060-5068.

Santos,M.A. and Tuite,M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res. *23*, 1481-1486.

Santos,M.A., Ueda,T., Watanabe,K., and Tuite,M.F. (1997). The non-standard genetic code of *Candida spp*.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423-431.

Schimmel,P. (2008). Development of tRNA synthetases and connection to genetic code and disease. Protein Science *17*, 1643-1652.

Schneider-Poetsch,T., Usui,T., Kaida,D., and Yoshida,M. (2010). Garbled messages and corrupted translations. Nat. Chem. Biol. *6*, 189-198.

Schramm,L. and Hernandez,N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev. *16*, 2593-2620.

Schultz,D.W. and Yarus,M. (1996). On malleability in the genetic code. J. Mol. Evol. *42*, 597-601.

Shah,P. and Gilchrist,M.A. (2010). Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. PLoS. Genet. *6*.

Soll,D. (2007). The genetic code revisited--four decades after Francis Crick. Nucleic Acids Symp. Ser. (Oxf) 13-14.

Soll,D. and RajBhandary,U.L. (2006). The genetic code - thawing the 'frozen accident'. J. Biosci. *31*, 459-463.

Soll,D.R. (2002). *Candida* commensalism and virulence: the evolution of phenotypic plasticity. Acta Trop. *81*, 101-110.

Soma,A., Kumagai,R., Nishikawa,K., and Himeno,H. (1996). The anticodon loop is a major identity determinant of *Saccharomyces cerevisiae* tRNA(Leu). J. Mol. Biol. *263*, 707-714.

Srinivasan,G., James,C.M., and Krzycki,J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Stansfield,I., Jones,K.M., Herbert,P., Lewendon,A., Shaw,W.V., and Tuite,M.F. (1998). Missense translation errors in *Saccharomyces cerevisiae*. J. Mol. Biol. *282*, 13-24.

Sudbery,P., Gow,N., and Berman,J. (2004). The distinct morphogenic states of *Candida albicans*. Trends Microbiol. *12*, 317-324.

Suzuki,T., Ueda,T., and Watanabe,K. (1997). The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122-1134.

Takase,Y., Takagi,T., Komarnitsky,P.B., and Buratowski,S. (2000). The essential interaction between yeast mRNA capping enzyme subunits is not required for triphosphatase function in vivo. Mol. Cell Biol. *20*, 9307-9316.

Thompson,D.M. and Parker,R. (2009). The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. J. Cell Biol. *185*, 43-50.

Trifonov,E.N. (2004). The triplet code from first principles. J. Biomol. Struct. Dyn. *22*, 1-11.

Weixlbaumer,A. and Murphy IV,F. (2009). Crystallographic Studies of Decoding by Modifi Ed Bases: Correlation of Structure and Function. In: DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution, ed. Henri GrosjeanAustin, Texas, USA: Landes Bioscience, 493-508.

Woese,C.R. (1965). On the evolution of the genetic code. Proc. Natl. Acad. Sci. U. S. A *54*, 1546-1552.

Wong,J.T. (1983). Membership mutation of the genetic code: loss of fitness by tryptophan. Proc. Natl. Acad. Sci. U. S. A *80*, 6303-6306.

Wong,J.T. (2005). Coevolution theory of the genetic code at age thirty. Bioessays *27*, 416-425.

Wong,J.T. (2007). Question 6: coevolution theory of the genetic code: a proven theory. Orig. Life Evol. Biosph. *37*, 403-408.

Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y., and Osawa,S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. U. S. A *82*, 2306-2309.

Yarus,M., Caporaso,J.G., and Knight,R. (2005). Origins of the genetic code: the escaped triplet theory. Annu. Rev. Biochem. *74*, 179-198.

Yuan,J., O'Donoghue,P., Ambrogelly,A., Gundllapalli,S., Sherrer,R.L., Palioura,S., Simonovic,M., and Soll,D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. FEBS Lett. *584*, 342-349.

Zaher,H.S. and Green,R. (2009). Fidelity at the molecular level: lessons from protein synthesis. Cell *136*, 746-762.

2. Variable genetic code ambiguity in *Candida albicans*

2.1. Abstract

The ascomycete *Candida albicans* is a successful commensal and a pathogen that occurs in a broad range of sites of the human body and other warm-blooded animals, demonstrating high capacity to survive and proliferate in environments with drastic changes in oxygen, carbon dioxide, pH, osmolarity, nutrients and temperature (Gasch, 2007). This opportunistic pathogen and other *Candida* spp. are characterized by the reassignment of the leucine CUG codon to serine (Santos *et al.*, 1993). Surprisingly, the serine tRNA ($tRNA_{CAG}^{Ser}$) that decodes CUG codons is recognized by both the leucyl - and the seryl-tRNA synthetase, allowing for incorporation of leucine and serine at CUG positions *in vivo*. The Seryl-tRNA synthetase (SerRS) gene (CaSES1) has one CUG codon at position 197 (Ser197) who's ambiguous decoding results in the production of two SerRS isoforms. Previous studies showed that the SerRS_Leu197 is more active although less stable than the SerRs_Ser197 and suggested the existence of a negative feedback loop to maintain the levels of Ser-$tRNA_{CAG}^{Ser}$ under strict control. In order to confirm the existence of such mechanism in *C. albicans in vivo* we have engineered strains expressing only one of the SerRS isoforms. The SerRS_Ser197 isoform recognized the $tRNA_{CAG}$ more efficiently than the SerRS_Leu197 isoform and outcompeted the LeuRS, reducing the level of the Leu-$tRNA_{CAG}^{Ser}$ in the cell. This observation questions the previous data and prompted us to investigate whether leucine incorporation at CUG positions fluctuates with physiology. To shed new light in this phenomenon, we engineered a fluorescent reporter system to monitor leucine incorporation at CUG codons in different physiological conditions. Our data show dynamic leucine incorporation at CUGs and dramatic increase in leucine incorporation levels in presence of macrophages and amphotericin B.

2.2. Introduction

The genetic code defines the rules that cell use to translate gene information into protein information. Two important classes of molecules involved

in this process are tRNAs and aminoacyl-tRNA synthetases (aaRS). There are 20 aaRS responsible for loading the correct amino acid into 20 tRNA families. This is highly specific biochemical reaction which is controlled by two types of identity elements namely identity determinants and antideterminants (Giege *et al.*, 1998). These identity elements occur mainly in the acceptor stem, in particular at the discriminator base (N73), the three first base pairs (1:72, 2:71 and 3:70), the three bases of the anticodon and the D loop (Giege and Lapointe, 2009). Members of aaRS class I approach the acceptor stem from the minor groove side of the tRNA while class II aaRSs approach it from the major groove side (Ribas de Pouplana and Schimmel, 2001)

In *C. albicans* the $tRNA_{CAG}^{Ser}$ has identity elements that are recognized by the SerRS namely the discriminator base (G73) and three G-C pairs in the variable arm and identity elements that are recognized by the LeuRS, namely A35 and m^1G37 in the anticodon loop and the A73 discriminator base (Soma *et al.*, 1996). Another interesting feature of the $tRNA_{CAG}^{Ser}$ is a guanosine at position 33 (G33) instead of the critical uridine (U33), which is required for the functional conformation of the anticodon U-turn and for correct stacking of the anticodon bases (Ashraf *et al.*, 1999). This G33 mutation lowers decoding efficiency and leucylation of the $tRNA_{CAG}^{Ser}$ (Santos *et al.*, 1996;Suzuki *et al.*, 1997). The atypical dual recognition of the *C. albicans* $tRNA_{CAG}^{Ser}$ by the LeuRS and SerRS allows for flexible leucine incorporation at CUG codons on a proteome wide scale. Previous studies carried out using high-pressure liquid chromatography (HPLC) mass spectrometry, showed that leucine misincorporation at CUG sites at 30°C, 37°C, in pH 4.0 and in presence of 1.5 mmol/l hydrogen peroxide (H_2O_2) are 2.96% , 3.9% , 4.95% to 4.03%, respectively (Gomes *et al.*, 2007).

In order to clarify whether CUG ambiguity varies during infection we have developed a green fluorescent reporter system (GFP) to quantify leucine insertion at CUG positions *in vivo*. Several conditions with biological significance, namely different pH, oxidative stress (H_2O_2), antifungal (amphotericin B, caspofungin), macrophage exposure and osmotic pressure (sorbitol) were monitored. The levels of leucine incorporation observed in presence of 0.40 µg/ml and 0.80 µg/ml amphotericin B were 8.7% and 10.1%, respectively. The other antifungal,

caspofungin, revealed values of 4.3% and 15.6%, for concentrations of 0.08µg/ml and 0.24 µg/ml, respectively. In 0.2M, 0.6M and 1M sorbitol, cells incorporated 9.2%, 6.1% and 15.7% of leucine at CUG sites, respectively. For H_2O_2 at 0.5 mM and 1 mM, leucine incorporation was 7.8% and 5.0%, respectively. Exposure to macrophages resulted in 3.98% of leucine incorporation, but simultaneous exposure to macrophages and amphotericin B increased leucine incorporation to 33.11% after 1h30m of exposure and to 49.33% after 3h of exposure.

These results invalidate the existence of a negative feedback mechanism to regulate leucine incorporation at CUG sites and show that *C. albicans* responds to stress by increasing leucine incorporation into the proteome.

2.3. Material and Methods:

2.3.1. Strain maintenance

The *C*. *albicans* strain SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434})(Noble and Johnson, 2005) was kind gift from Alexander D. Johnson (University of California-San Francisco). SN148 was grown in YPD (1% yeast extract, 2% peptone, and 2% dextrose) at 30ºC. Transformed strains were grown in synthetic minimal medium (0.67% yeast nitrogen base, 2% glucose, 0.2% Drop-out mix with all the essential amino acids). *E. coli* JM109 (*recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, Δ(*lac-proAB*)/F' [*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15]) was grown in LB (1% peptone from casein, 0.5% yeast extract, 1% sodium chloride) at 37ºC., when necessary LB was supplemented with ampicillin (75 mg/ml).

2.3.2. Plasmid purification and construction

The plasmids used in this study were purified using GeneJet^{TM} Plasmid Miniprep Kit (Fermentas) according to the manufacturer's instructions. Plasmid CIp20 (Dennison *et al.*, 2005) was digested with *SacI* and *SpeI* resulting in the

deletion of the *Ura3* marker gene. *C. albicans SerRS* gene was amplified by PCR with the forward primer oUA1640 5'-TTTGAGCTCCCCATAGTTTGATAGACGAAT -3' and reverse oUA1642 5'-TTTACTAGTTTTGGTGGAGGTTGAGATAC -3', and was introduced between *SacI* and *SpeI* restriction sites. Then this plasmid was digested with *BamHI* and *SalI* and the *RPS1* was removed, the *SerRS* was then amplified with forward primer oUA 1643 5'-ATATGTACCTCGACAGTCATCGGATCCTTT-3' and reverse primer oUA1644 5'-TTTGTCGACCAATCAAATTTTTACTGCTTC-3' and inserted between the *BamHI* and *SalI* restriction sites, resulting in plasmid pUA520. The *C. albicans SerRS* gene was amplified with the forward primer oUA1640 and reverse primer oUA1641 5'- AAAGCGGCCGCTTTGGTGGAGGTTGAGATAC -3', and was cloned between the *SacI* and *NheI* restriction sites of CIp30 (Dennison *et al.*, 2005) plasmid. Next the *RPS1* was removed, by digesting the plasmid with *BamHI* and the SalI, then the *SerRS* genes was amplified with forward primer oUA1643 and reverse primer oUA1644 and inserted between the *BamHI* and *SalI* restriction site, resulting in plasmid pUA523. To construct the SerRS isoform containing a Leucine at codon_197, the pUA520 and pUA523 plasmids were mutated to change 5'-CTG-3' to 5'-TTA-3'(Leu) using the QuikChange Site-Directed Mutagenesis kit (Stratagene) with the forward primer oUA1576 5'-GATCAACTACGGTTTATTATTTTTGAGTAGCAAAGG-3' and reverse primer oUA1577 5'- CCTTTGCTACTCAAAAATAATAAACCGTAGTTGATC-3', originating plasmids pUA521 and pUA524, respectively. The plasmids containing the SerRS isoform with Ser_197 was constructed by mutating plasmids pUA520 and pUA523 with the forward primer oUA1578 5'-GATCAACTACGGTTTATCTTTTTTGAGTAGCAAAGG-3' and reverse primer oUA1579 5'- CCTTTGCTACTCAAAAAAGATAAACCGTAGTTGATC-3' resulting in the plasmids pUA522 and pUA525, respectively.

The plasmids used in this study to quantify Leucine incorporation in *C. albicans in vivo*, were based on the pACT1–GFP (Barelle *et al.*, 2004a), with the following modifications. *LEU2* from *Candida maltosa* was amplified by PCR from pSN40 (Noble and Johnson, 2005) with primers oUA1820 5'-

TTTACTAGTGGATCCAATCATCACTGGTG-3' and oUA1821 5'-TTTGCGGCCGCACCTACCCATGTCTAGAAAG-3', and was inserted between *SpeI* and *NotI* restriction enzyme sites, replacing the *Ura3* gene (resulting in plasmid pUA556). The codon_201 of yEGFP in plasmid pUA556 was mutated from 5'-TTA-3' to 5'-CTG-3' with the forward primer oUA1576 5'-GTTACCAGACAACCATTACCTGTCCACTCAATCTGCCTTATC-3'and reverse primer oUA1577 5'-GATAAGGCAGATTGAGTGGACAGGTAATGGTTGTCTGGTAAC-3', originating plasmid pUA557. Plasmid pUA558 was produced by mutating the codon_201 of yEGFP in plasmid pUA556 from 5'-TTA-3' to 5'-TCT-3' with the forward primer oUA1578 5´-CCAGACAACCATTACTCTTCCACTCAATCTGCC-3´ and the reverse primer oUA1579 5´-GGCAGATTGAGTGGAAGAGTAATGGTTGTCTGG-3.

The plasmid used to quantify leucine incorporation at CUG codons in SN148 strains was constructed as described below. The mCherry gene encoded in plasmid pMG2288 was amplified in two rounds to add a His-Tag, the first round used primers oUA1813 5'-AAAGGATCCCTCGAGGAGCTATTAAGATCACCA-3', oUA1814 5'-GTCGTCATCCTTGTAATCTTTATATAATTCATCCATACCACC-3', the second round was performed with primers oUA1813 and oUA1815 5'-AAAACTAGTTTACTTATCGTCGTCATCCTTGTAATCTTTATA-3'. The PCR product was then inserted into plasmid pACT1-GFP (Barelle *et al.*, 2004b) between restriction sites *BamHI* and *SpeI*, originating plasmid pUA553. The codon_201 of yEGFP in plasmid pUA553 was mutated from 5'-TTA-3' to 5'-CTG-3' using the QuikChange Site-Directed Mutagenesis kit (Stratagene) with the forward primer 5'-GTTACCAGACAACCATTACCTGTCCACTCAATCTGCCTTATC-3' and the reverse primer 5'-GATAAGGCAGATTGAGTGGACAGGTAATGGTTGTCTGGTAAC-3', resulting in plasmid pUA554. Another plasmid (pUA555) was also constructed by mutating the codon_201 of yEGFP from 5'-TTA-3' to 5'-TCT-3' using the QuikChange with the forward primer 5´-CCAGACAACCATTACTCTTCCACTCAATCTGCC-3´ and reverse primer 5´-GGCAGATTGAGTGGAAGAGTAATGGTTGTCTGG-3.

2.3.3. Construction of strains

C. albicans expressing single SerRS isoforms:

The protocol used to construct *C. albicans* strains expressing a single isoform of the SerRS protein was based on a gene disruption protocol described previously (Noble and Johnson, 2005). Briefly, plasmids pUA521, and pUA522 were digested with *SacI* and *SalI* and transformed in SN148 *C. albicans* cells, and selected in minimal medium lacking histidine. The corrected substitution of SerRS was verified by PCR amplification with primers oUA1650 5'-GTTTTACCTTTATGTCCCAG-3' and oUA1635 5'-TGTGACTTGGGTATAACTGG-3'. Then plasmids pUA524 and pUA525 were digested with *SacI* and *SalI*, and transformed in the SN148 *C. albicans* cells transformed previously with pUA521 and pUA522, respectively. The authenticity of the insertions was confirmed by PCR amplification using primers oUA1650 and oUA1610 5'- GTACACGACCCACAGTTAG1TC-3'.

Construction of strains to measure leucine incorporation at CUG sites

Plasmids pUA553, pUA554 and pUA555 were linearized with *StuI* and transformed in *C. albicans* SN148 strain. Correct integration at the *RPS10* locus was confirmed by PCR using the primers oUA1554 5'-CGTATTCACTTAATCCCACAC-3', oUA1555 5'-CCAATTGGTGATGGTCC-3'. Single copy integrant were selected using the primers oUA1556 5'-GGTATAGAAATGCTGGTTGG-3' and oUA1557 5'-CCAATTGGTGATGGTCC-3', which only give a product for tandem integrated plasmids (Barelle *et al.*, 2004b).

2.3.4. *C. albicans* transformation

Transformation of *C. albicans* was carried out using an improved lithium acetate method (Walther and Wendland, 2003) with small modifications. Briefly overnight cultures were inoculated in fresh medium to an $O.D_{600nm}$ of 0.3 and

allowed to grow at 30ºC, 200rpm shaking until an $O.D_{600nm}$ of 1-1.2 was reached. Cell cultures were then transferred to 50ml falcon tubes and centrifuge 4000rpm for 5min, supernatants were discarded and pellets were resuspended in 1.5ml of LiAc-solution (0.1M LiAc, TE buffer 1x). 200 µl of suspension was transferred to 1.5mL eppendorf tubes and the transformation components were added in the following order: 600µL 50% (w/w) PEG LiAc-solution, 50µL single-stranded carrier DNA (2mg/mL) previously denatured and 50µL of an aqueous solution of the plasmid of interest (containing 1– 5µg of plasmid). Tubes were vortexed and incubated at 30ºC for 4 hours following incubation at 44ºC for 15min and another one in ice for 2min. Cells were harvested at 4000rpm for 5min and resuspended gently in 200µl of appropriate SC selective medium. Aliquots of 100µl of cell suspensions were plated onto appropriate SC selective medium plates and incubated at 30°C for 3-4 days.

2.3.5. Growth curves

C. albicans transformed strains were grown until late stationary phase at 30ºC, 180 rpm. Next, strains were inoculated to an initial $O.D._{600nm}$ of 0.02 in 100 ml Erlenmeyer flasks. At several time points the $O.D._{600nm}$ of the culture was measured. The growth rate was calculated using cultures growing exponentially.

2.3.6. Protein extraction

Protein extraction was prepared essentially as described by Rand and Grant (Rand and Grant, 2006) with minor alterations. Cells were grown at 30°C, 200 rpm, until and $O.D._{600nm}$ of 0.4 – 0.5, and then harvested at 4000rpm, 4°C, for 5 min. At this point pellets were washed with 1 ml of sterile cold water, transferred to a 2 ml microtube and frozen at -80ºC for 30min at least. Afterwards, 300µl of lysis buffer [50 mM potassium phosphate buffer, pH 7, 1 mM EDTA, 5% (vol/vol) glycerol, 1 mM phenylmethylsulfonyl fluoride, and Complete Mini protease inhibitor cocktail (Roche)] was added to each tube. Cells were disrupted with 100µl of glass beads using a Precellys 24 (bertin technologies) 3X at 5000 rpm for 10 s,

followed by 2min of incubation on ice. The mixture was centrifuged at 3000g for 15 min at 4ºC and supernatants were transferred to a new tube, obtaining total protein. Protein quantification was determined using the Pierce® BCA Protein Assay Kit (Thermo SCIENTIFIC), according to the manufacturer´s instructions.

2.3.7. Western blot analysis

For routine western blotting 30 µg of total protein extract were fractionated on a 15% SDS. GFP was detected using a mouse anti-GFP primary antibody (Clontech, 1:1000) for 2 hours and a secondary antibody IRDye600 Goat anti-mouse (Li-cor Biosciences, 1:10000), incubated for 1 hour. For sample loading normalization, membranes were stripped and re-incubated with a rabbit anti-ADH antibody (1:1000) for 2 hours followed by detection with IRDye600 Goat anti-rabbit secondary antibody (Li-Cor Biosciences), incubated for 1 hour. Detection was performed using the ODYSSEY Infrared Imaging System (Li-Cor Biosciences) with the adequate definitions.

2.3.8. Quantification of leucine incorporation at CUG sites

For leucine quantification the set of strains constructed carry an integrated plasmid expressing mCherry and GFP. The positive control strain carries recombinant GFP with a leucine codon TTA in position 201, while the negative control strain carries a serine codon TCT at position 201. In the reporter strain the codon at position 201 is the ambiguous CTG codon.

2.3.9. Stress conditions

Strains were grown in liquid media with compound concentrations indicated in Table 2-1, until $O.D_{600nm}$ of 0.5. Samples were then collected and observed under the microscope.

Table 2-1: Conditions used to determine leucine incorporation at CUG sites.

Assay	Stress compound	Concentration	Base medium	Growth temperature
Control	pH 7.0		MM-Uri	30ºC
pH value	pH 4.0		MM-Uri	30ºC
	pH 5.0		MM-Uri	30ºC
	pH 6.0		MM-Uri	30ºC
	pH 7.0		MM-Uri	30ºC
	pH 8.0		MM-Uri	30ºC
Oxidative stress	Hydrogen Peroxyde	0.50 mM	MM-Uri	30ºC
		1.00 mM	MM-Uri	30ºC
Antifungal	Amphotericin B	0.40 µg/ml	MM-Uri	30ºC
		0.80 µg/ml	MM-Uri	30ºC
	Caspofugin	0.08 µg/ml	MM-Uri	30ºC
		0.24 µg/ml	MM-Uri	30ºC
Osmotic Stress	Sorbitol	0.20 M	MM-Uri	30ºC
		0.60 M	MM-Uri	30ºC
		1.00 M	MM-Uri	30ºC

2.3.10. Macrophage Tests

C. albicans strains were grown on solid MM-Uri incubated at 30 °C for at least 2 days, and then collected and diluted to the desired cell dilution in PBS. Cells were then added in equal proportion to macrophages maintained in complete culture medium (RPMI1640 medium) with the desired antifungal concentration. Culture plates were incubated between 1h30m and 3h at 37 °C in a humidified atmosphere with 5% CO_2. Macrophages were washed to select those bound to the cover slip containing intracellular *C. albicans* cells. Samples were fixed in 4% Paraformaldehyde and fluorescent images of cells were obtained using a 40X objective and an epifluorescence up-right microscope Imager.Z1 (Zeiss), equipped with 38 HE GFP filter, a 63 HE mRFP, AxioCam HRm camera (Zeiss) and AxioVision software (Zeiss).

2.3.11. Microscopy

The expression of mCherry and GFP was monitored in *C. albicans* cells using epifluorescence microscopy. Fluorescent images of cells were obtained using a 40X objective in an epifluorescence up-right microscope Imager.Z1 (Zeiss), equipped with 38 HE GFP filter, a 63 HE mRFP, AxioCam HRm camera

(Zeiss) and AxioVision software (Zeiss). Single cell fluorescence intensity quantification was carried out using the ImageJ software package.

Leucine incorporation was quantified using each GFP intensity value, which was divided by mCherry intensity values obtained for the same cell. Fluorescence values obtained and average for the negative control strain were subtracted from the values obtained for the reporter and positive control strains. Finally leucine incorporation was calculated by dividing the fluorescence values obtained for the reporter strain by the values obtained for the positive control strain average value (Equation 1).

- $GFP_{CUG(Reporter)}$= GFP intensity values of cells with plasmid pUA554.
- $mCherry_{(Reporter)}$= mCherry intensity values of cells with plasmid pUA554.
- $GFP_{UCU(Negative)}$= GFP intensity values of cells with plasmid pUA555.
- $mCherry_{(Negative)}$= mCherry intensity values of cells with plasmid pUA555
- $GFP_{UUA(Positive)}$= GFP intensity values of cells with plasmid pUA553
- $mCherry_{(Positive)}$= mCherry intensity values of cells with plasmid pUA553

$$\text{Leucine incorparition at CUG codons} = \frac{[GFP_{CUG\,(Reporter)}/mCherry_{(Reporter)}] - [GFP_{UCU\,(Negative)}/mCherry_{(Negative)}]}{[GFP_{UUA\,(Positive)}/mCherry_{(Positive)}] - [GFP_{UCU\,(Negative)}/mCherry_{(Negative)}]}$$

Equation 1: Leucine incorporation at CUG codons.

2.4. Results

2.4.1. SerRS

The SerRS is responsible for loading serine on $tRNAs^{Ser}$. The *C. albicans* SerRS gene has one CUG codon at position 197 and subsequently two SerRS isoforms are produced in *C. albicans* cells. Previous studies revealed that the SerRS_Leu197 is more active although less stable than SerRs_Ser197, suggesting the existence of a negative feedback mechanism to maintain CUG ambiguity at low level (Rocha *et al.*, 2011b). The suggested mechanism relies on increased misincorporation of Leu at CUG codons, which consequently increases

the levels of the apparently more active SerRS_Leu197 isoform resulting in increased serylation of the $tRNA_{CAG}^{Ser}$. The SerRS_Leu197 competes more efficiently with the LeuRS for charging of the $tRNA_{CAG}^{Ser}$, thus decreasing the level of the misincorporation at CUGs (Rocha *et al.*, 2011b). Since this feedback mechanism was proposed on the basis of in *in vitro* biochemical data and also on the orthogonal transplant of the SerRS- $tRNA_{CAG}^{Ser}$ pair into *E. coli*, we engineered *C. albicans* strains to express one isoform of the SerRS only (Figure 2-1) and quantified the level of CUG ambiguity in each type of cells.

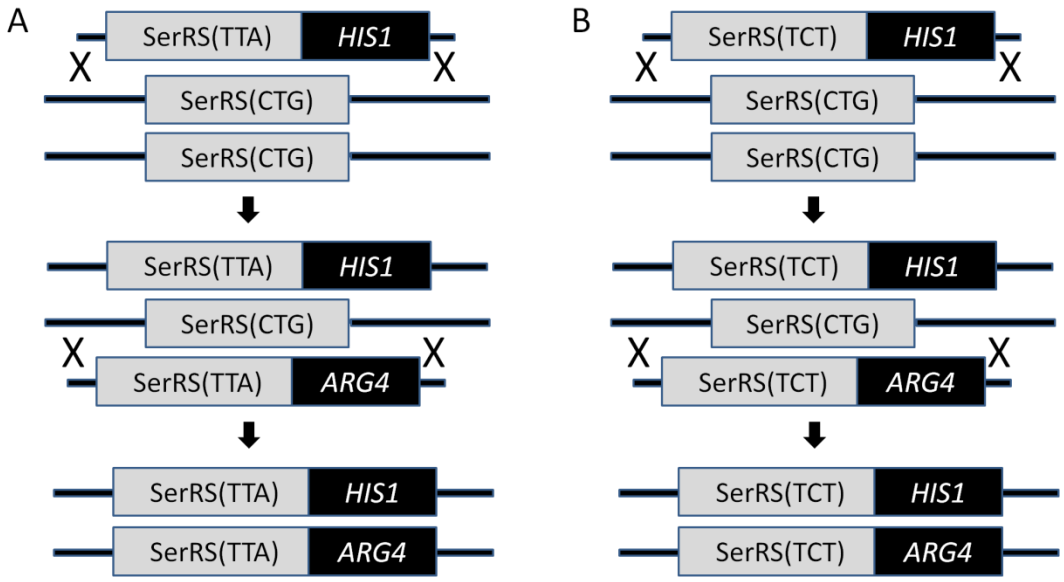


Figure 2-1: Engineering strains of *C. albicans* to express one of the two SerRS isoforms only. The SerRS gene substitution was performed as described by Noble and Johnson (Noble and Johnson, 2005). A) The SerRS gene encoding a CTG codon at position 197 was substituted by a SerRS gene with a leucine TTA codon at position 197, the same process was repeated in the other allele. B) The SerRS gene with a codon CTG at position 197 was substituted by a SerRS gene containing a serine TCT codon at position 197, the same process was repeated in the other allele.

The *C. albicans* strains expressing only one SerRS isoform, SerRS_Leu197 or SerRs_Ser197, showed similar growth rate of $0.30h^{-1}$ and $0.29h^{-1}$, respectively (Figure 2-2), suggesting that both SerRS isoforms aminoacylate the serine tRNAs with similar efficiency.

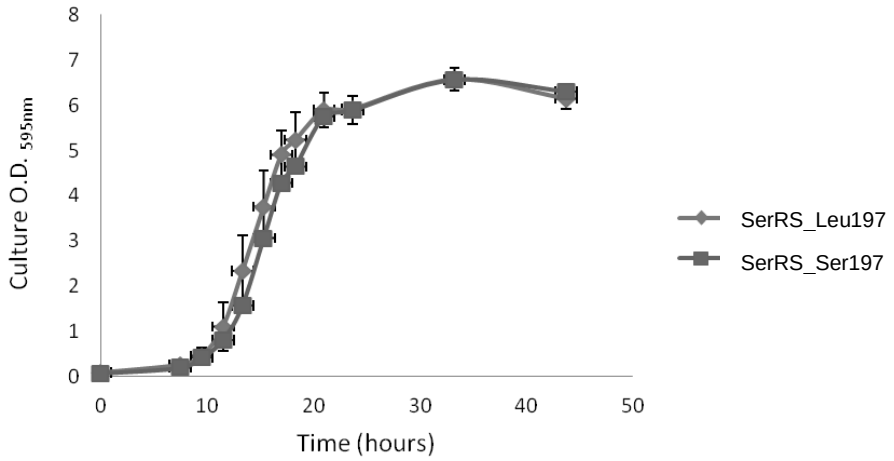


Figure 2-2: Growth curve of *C. albicans* strains expressing one of the two SerRS isoforms, SerRS_Ser197 or SerRS_Leu197.

In order to quantify the level of leucine incorporation at CUG codons in cells expressing the SerRS_Ser197 and in cells expressing the SerRS_Leu197, we used a reporter system based on the yeast-enhanced GFP, whose open reading frame (orf) contains optimal codons for *C. albicans* (Cormack *et al.*, 1997). Our reporter system includes a positive control (pUA556), which corresponds to synthesis of stable GFP and a CUG reporter (pUA557) which contains the CUG codon at position 201, a position that was previously described as being essential for GFP stability (Cormack *et al.*, 1997). This reporter only produces stable GFP when leucine is incorporated at position 201. A negative control (pUA558) that has a TCT serine codon at position 201 and produces inactive GFP was also constructed. The expression of GFP was quantified by western blot analysis (Figure 2-3). The synthesis of stable GFP (leucine incorporation at position 201) was higher in the strain expressing the SerRS_Leu197 than SerRS_Ser197 isoform, about 3.97% and 0.87%, respectively. In other words, this result contradicts the *in vitro* biochemical data which showed that the SerRS_Leu197 is more active than the SerRS_Ser197 (Rocha *et al.*, 2011b), because increased levels of leucine incorporation at position 201 of GFP indicate that the LeuRS out competes more efficiently the SerRS_Leu197 than the SerRS_Ser197.

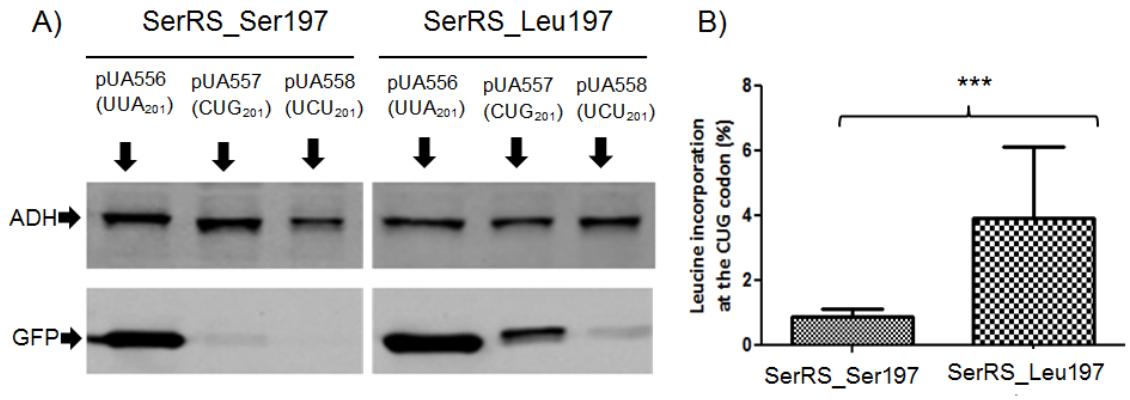


Figure 2-3- Quantification of functional GFP produced in *C. albicans* strains expressing only one of the two SerRS isoforms. A) Western blot analysis SerRS showing that *C. albicans* cells expressing the SerRS_Leu197 isoform incorporate higher levels of leucine at CUG codons. 30 µg of total protein extracts were separated by SDS-PAGE. **B)** Quantitative analysis of leucine incorporation at CUG codons. Data represents mean ± s.d.. Statistical analysis was performed using a t test (***p<0.001). pUA556- positive control that produces stable GFP, pUA557- reporter which contains the CUG codon at position 201 that produces stable GFP when leucine is incorporated at position 201, pUA558- negative control that has a TCT serine codon at position 201 and produces inactive GFP.

2.4.2. Quantification of Leucine incorporation at CUG codons

The lack of a negative feedback mechanism to control Leu incorporation at CUG codons prompted us to determine whether Leu/Ser ratios were sensitive to physiological conditions.

C. albicans colonizes several niches characterized by constant variation of nutrients, temperature, osmolarity, pH, toxins and also host defense mechanisms that involve reactive oxygen and nitrogen species generated by the immune system (Gasch, 2007). With the aim of monitoring misincorporation of leucine at CUG codons *in vivo* under these variable conditions, we have developed a fluorescent reporter system (Figure 2-4). Other reporter systems to monitor leucine incorporation have also been engineered, namely a reporter based on a heterologous protein containing a CUG codon, where the resulting peptides were identified and quantified using HPLC-mass spectrometry (Gomes *et al.*, 2007). But, such reporter systems do not permit quantifying leucine incorporation at the single cell level and are very complex and time consuming. Our reporter system includes two fluorescent proteins, GFP and mCherry and several controls, namely,

a strain harboring a positive control that produces stable GFP, a strain expressing a reporter GFP that encodes a CTG codon at position 201 and only produces stable GFP when leucine is incorporated at CTG-201, and a strain expressing a negative control gene which does not produce stable GFP. This strain has serine TCT_201 in the GFP gene and serine incorporation at this position destabilizes and targets GFP for degradation. These strains also express mCherry which is used to normalize protein levels. The quantification of GFP and mCherry was carried by epifluorescence microscopy with appropriate filter sets, using identical time exposures to capture images (Figure 2-5) and posterior image analysis using ImageJ.

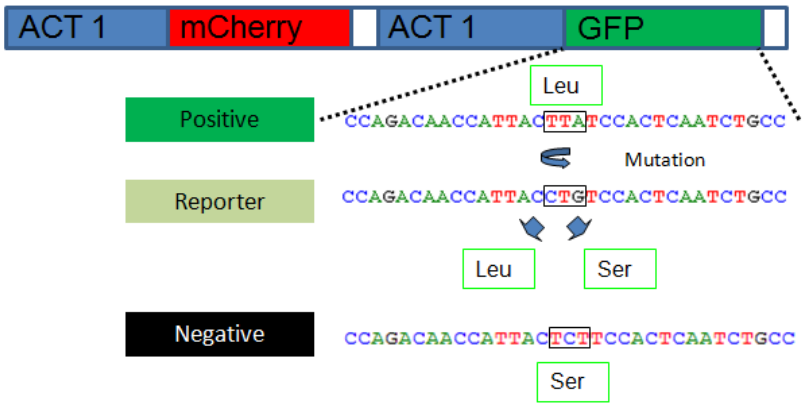


Figure 2-4: Scheme of the reporter system built to quantify leucine incorporation at CUG codons. The system is based on the plasmid pACT1-GFP, which contains the codon-optimized yeast enhanced green fluorescent protein (yEGFP) (Barelle et al 2004). The leucine-201 is encoded by the Leu_UUA codon which was mutated to the ambiguous CUG codon and also to Ser_UCU codon (negative control).

The above genes were integrated into the *C. albicans* strain SN148 at the RPS10 locus and cells were grown in different physiological conditions, namely the presence of the antifungals amphotericin and caspofungin, macrophages, sorbitol and several pH and oxidative stress conditions.

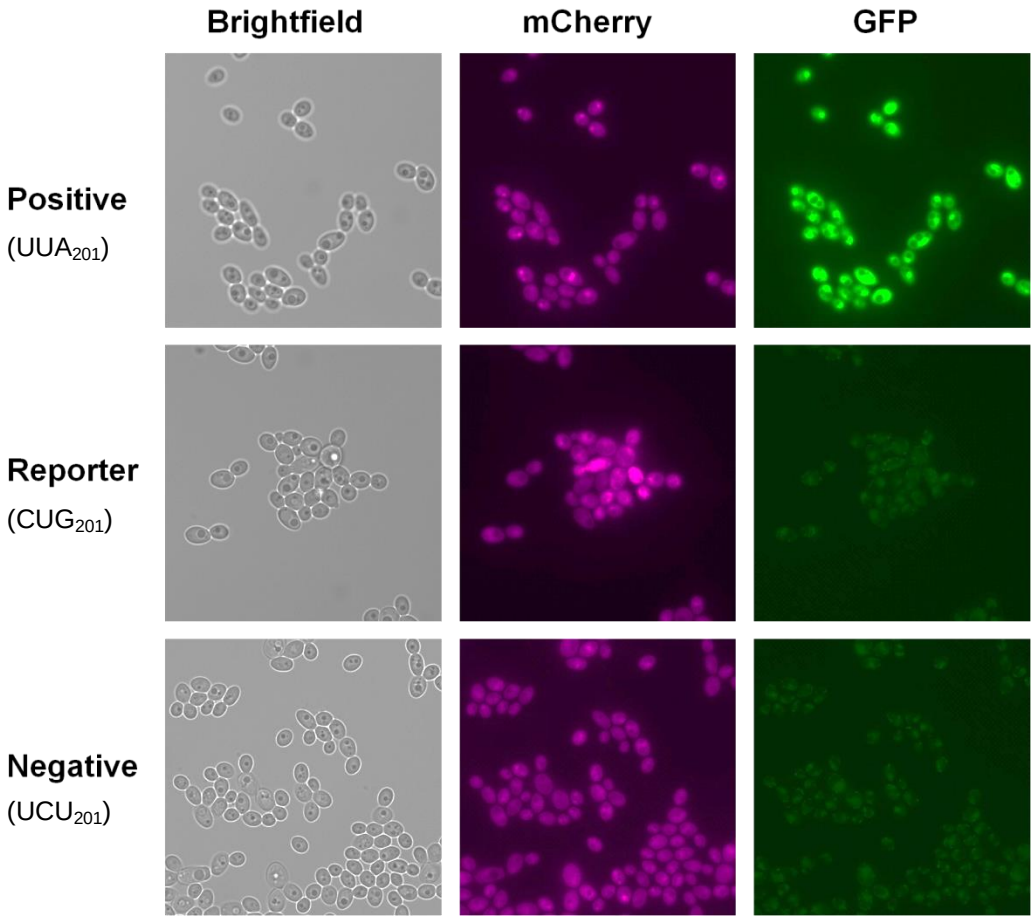


Figure 2-5: *In vivo* quantification of leucine incorporation at CUG codons. The positive control produces stable GFP. The Reporter GFP encodes CUG_201 codon and the negative control encodes UCU_Ser codon at position 201. All strains produce stable mCherry which is used to normalize protein levels. Fluorescence and brigthfield images were obtained using a 40X objective in an epifluorescence up-right microscope Imager.Z1 (Zeiss), equipped with 38 HE GFP filter, a 63 HE mRFP, AxioCam HRm camera (Zeiss) and AxioVision software (Zeiss).

The antifungal amphotericin B is a polyene that inserts into lipid bilayers, binds to sterols and forms pores in the cell membrane. These pores disrupt plasma membrane integrity permitting efflux of cations such as K^+. Polyenes also cause oxidative damage (Cannon *et al.*, 2007). Leucine incorporation at CUG_{201} in medium containing 0.40 µg/ml and 0.80 µg/ml of amphotericin B were 8.7% and 10.1%, respectively (Figure 2-6). Caspofungin is an antifungal that inhibits (1,3)-D-β-glucan synthase interfering with cell-wall biosynthesis (Cannon *et al.*, 2007). When *C. albicans* was exposed to this antifungal leucine incorporation at CUG_{201} was 4.3% and 15.6% for concentrations of 0.08 µg/ml and 0.24 µg/ml, respectively (Figure 2-6).

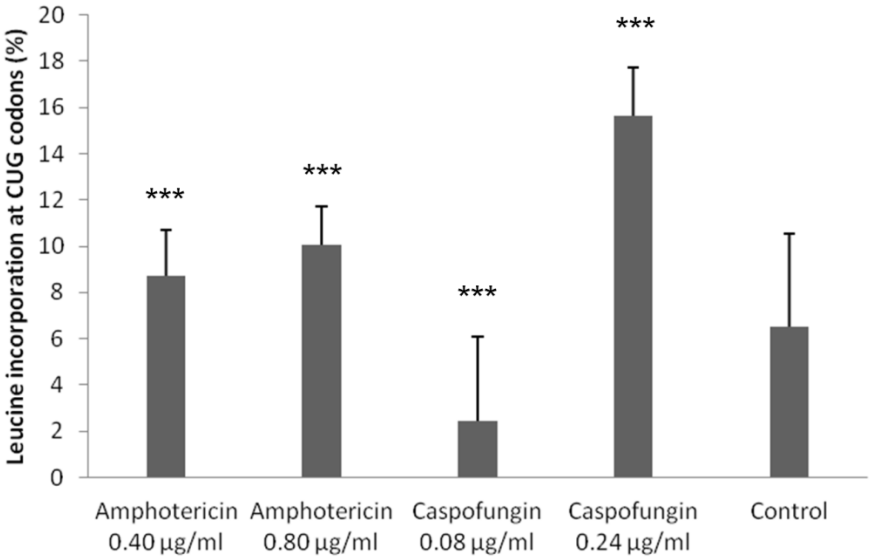


Figure 2-6: Leucine incorporation at GFP CUG_{201} in presence of antifungals. Data represents Leucine incorporation at GFP CUG_{201}, mean + s.d. of at least 100 cells of 2 different clones. Data statistical analysis was carried out using one-way ANOVA, followed by a Dunnet test with CI 95% relative to control (***p<0.001).

Several osmotic pressure conditions were tested using sorbitol, which is not assimilated by yeast cells (Hirasawa *et al.*, 2006). Sorbitol concentrations tested were 0.2 M, 0.6 M and 1 M and the leucine incorporation levels at GFP CUG_{201} were 9.2%, 6.1% and 15.7%, respectively (Figure 2-7).

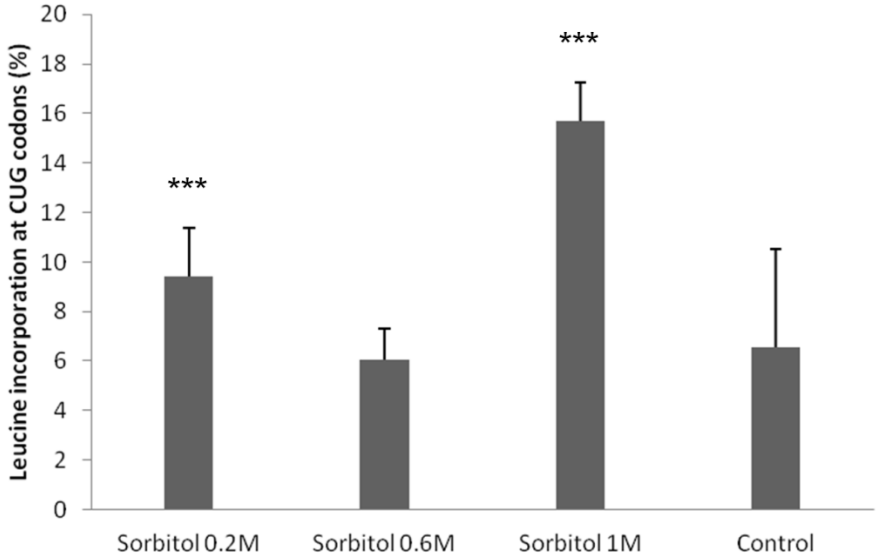


Figure 2-7: Leucine incorporation in GFP CUG_{201} site in *C. albicans* cells exposed to osmotic stress. Data represents Leucine incorporation ratio mean + s.d. of at least 100 cells of 2 different clones. Data statistical analysis was carried out using one-way ANOVA followed by a Dunnet test with CI 95% relative to control (***p<0.001).

The immune system responds to *C. albicans* invasion by producing antimicrobial compounds such as reactive oxygen species (ROS) (Chauhan *et al.*, 2006;Kumamoto, 2008). This prompted us to quantify leucine incorporation at CUG in presence of oxidative stress, namely in presence of 0.5mM and 1mM of hydrogen peroxide (H_2O_2). The level of leucine incorporated at the GFP CUG_{201} was 7.8% and 5.0%, respectively (Figure 2-8).

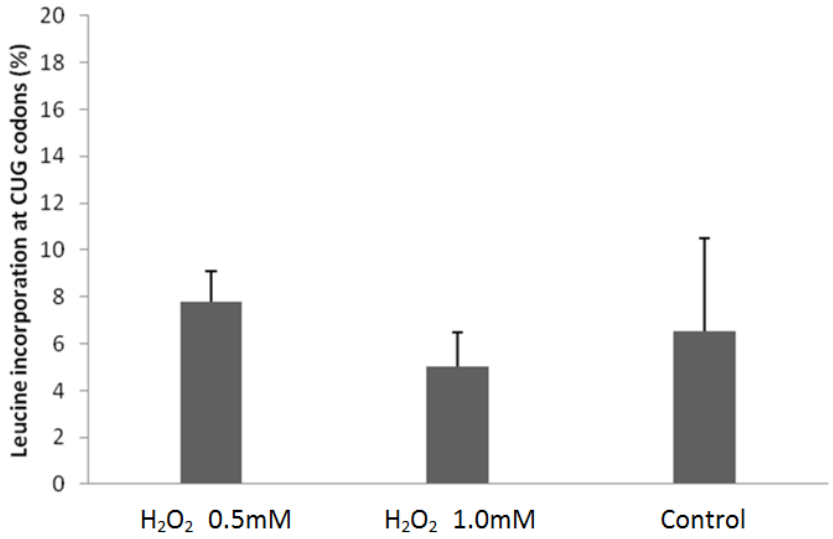


Figure 2-8: Leucine incorporation at CUG codons in *C. albicans* cells grown in presence of oxidative stress. Data represents the mean + s.d. of leucine incorporation in the GFP CUG_{201} of at least 100 cells of 2 different clones. Data statistical analysis was carried out by one-way ANOVA followed by a Dunnet test with CI 95% relative to control.

C. albicans is also able to colonize anatomical sites of diverse pH values, including the oral and gastro-intestinal tracts, the vaginal cavity and even bloodstream and cause systemic disease, where it can infect virtually all tissues of the host. Therefore, its capacity to adapt to changes in extracellular pH is probably important for its success as a commensal and as a pathogen. Remarkably, *C. albicans* can grow *in vitro* in an wide range of extracellular pH, between pH values of 2–10 (Davis, 2003). Several pH conditions were selected for leucine quantification (pH 4, 5, 6, 8) and the incorporation values obtained were 8.2%, 5.3%, 11.7% and 7.3%, respectively (Figure 2-9). The leucine incorporation levels observed at 30ºC were similar to those observed at pH 7 (6.5%).

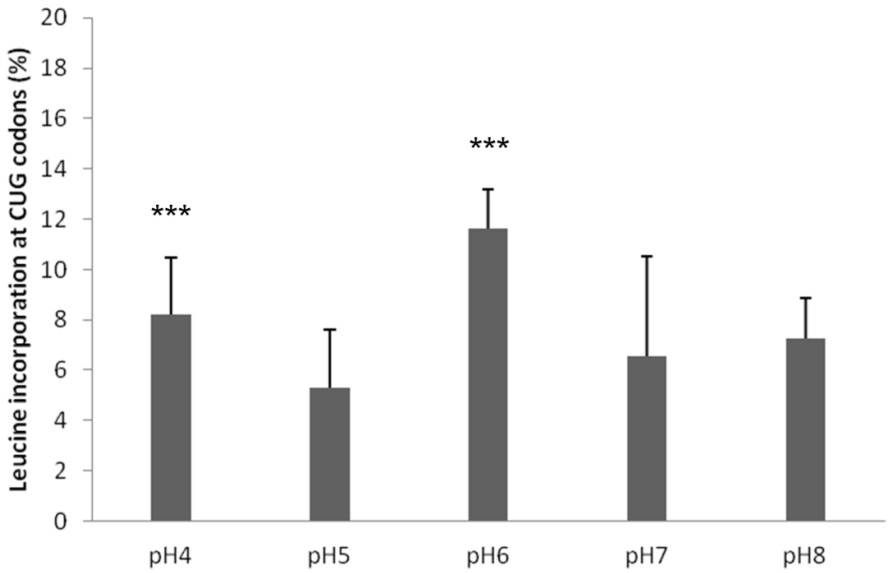


Figure 2-9: Leucine incorporation at GFP CUG_{201} in *C. albicans* cells grown in media with different pH values. Data represents mean + s.d. of leucine incorporation at GFP CUG_{201} of 100 cells of 2 different clones. Data statistical analysis was carried out using one-way ANOVA followed by a Dunnet test with CI 95% relative to pH 7 (***p<0.001).

The host defense against *C. albicans* infection depends mainly on the ingestion and elimination of invasive *C. albicans* by the innate immune system, mainly by macrophages and neutrophils (Lewis *et al.*, 2012). We have also determined the level of leucine incorporation at CUG codons in presence of macrophages and in presence of a combination of macrophages and antifungals. For this *C. albicans* cells were exposed to macrophage for 1h30m and 3h in the presence of the antifungal amphotericin B. The leucine incorporation at GFP CUG_{201} in presence of macrophages was 3.98%, but when exposed to the combination of macrophages and amphotericin B we observed a sharp increase in incorporation up to 33.11% after 1h30m of exposure and 49.33% after 3h of interaction (Figure 2-10).

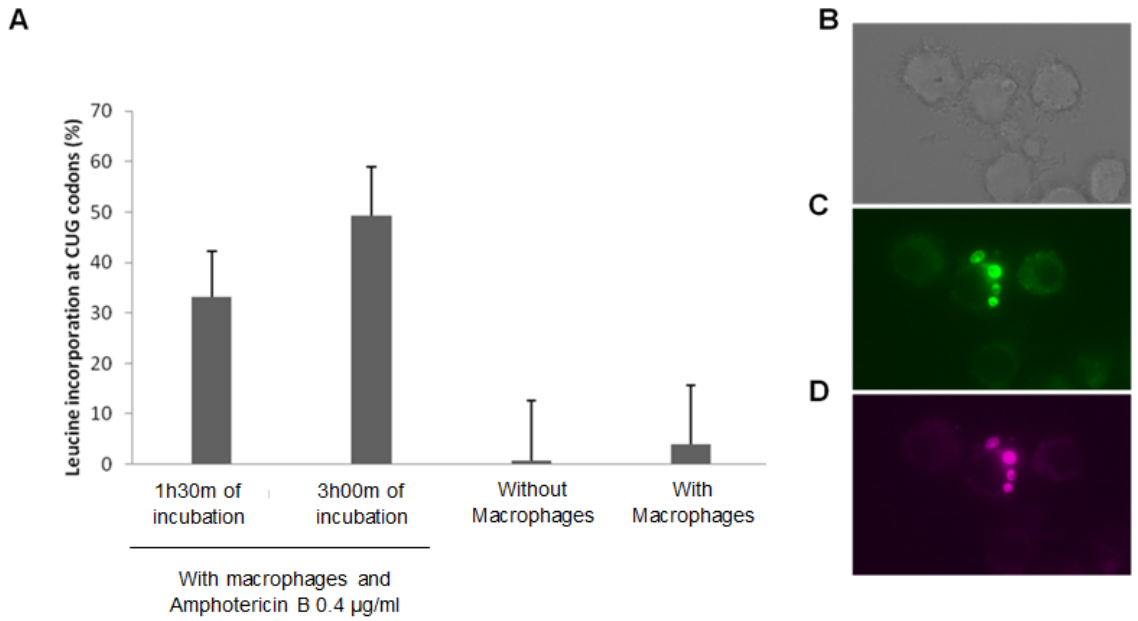


Figure 2-10: Leucine incorporation at CUGs in presence of macrophages and antifungals. A) Data represents mean ± s.d. of leucine incorporation. Macrophages and *C. albicans* interaction in Brigthfield (**B**) GFP (**C**) and mCherry (**D**).(Fluorescence and brigthfield photos were acquired using a 40X objective and a epifluorescence up-right microscope Imager.Z1 (Zeiss), equipped with 38 HE GFP filter, a 63 HE mRFP, AxioCam HRm camera (Zeiss) and AxioVision software (Zeiss).

2.5. Discussion

In this study we have constructed *C. albicans* strains that express one of the two SerRS isoforms only. These strains showed similar growth rate, although leucine incorporation was higher in the strain expressing the SerRS_Leu197 isoform. These results are in agreement with previous data showing that *C. albicans* can misincorparate 28% of leucine at CUG codons without visible effects on fitness (Gomes *et al.*, 2007;Miranda *et al.*, 2007). Although CUG codons are present in 67% (4168) of *C. albicans* protein-coding genes, 90% of these CUG codons are located at positions where both aliphatic leucine or polar serine can be inserted without causing disruption of protein structure and function (Rocha *et al.*, 2011b). Therefore, the capacity to tolerate high and fluctuating leucine misincorporation at CUG codons results from strategic locations of CUG codons in *C. albicans* protein coding genes (Rocha *et al.*, 2011b). The CUG codons where this ambiguous translation may lead to synthesis of proteins with different

properties are located in genes involved with virulence and pathogenesis, namely, biofilm formation, morphogenesis, and mating (Rocha *et al.*, 2011b), indicating a possible role for ambiguity in adaptation and infection.

Our data also show the *C. albicans* strains that express the SerRS_Leu197 isoform incorporate higher levels of leucine at CUG codons (3.97%) than the strains that express the SerRS_Ser197 (0.87). Therefore the SerRS_Leu197 is less active than the SerRS_Ser197. These results do not support the negative feedback mechanism proposed previously (Rocha *et al.*, 2011b). In other words, the higher levels of leucine incorporation at the GFP CUG_{201} can only be explained if the LeuRS outcompetes the SerRS_Leu197 more efficiently than the SerRS_Ser197 isoform. Several reasons may be pointed for the divergent results. *In vitro* SerRS activity was performed at 25ºC (Rocha *et al.*, 2011b) but the optimal grow temperature of *C. albicans* is 37ºC*,* and the SerRS_Leu197 isoform is less thermal stable between 30-45ºC (Rocha *et al.*, 2011a).

Our data also unveiled major fluctuation in the level of leucine incorporation at CUGs in different physiological conditions, raising the question of how to explain such alterations in leucine incorporation. One possibility is that the expression of the SerRS decreases under stress, while expression of the LeuRS is stress insensitive. Indeed, previous quantitative analysis showed that the cellular abundance of the SerRS decreases in presence of macrophage (Fernandez-Arenas *et al.*, 2007). This decrease of SerRS levels likely results in lower relative ratio of SerRS/LeuRS, increasing $tRNA_{CAG}^{Ser}$ aminoacylation with leucine, due to the ability of the LeuRS to recognize this tRNA.

The levels of leucine incorporation at CUG positions using our fluorescent reporter system were higher than the levels obtained previously using mass-spectrometry. For example, Gomes and colleagues detected 2.96% and 4.95% of leucine incorporation at 30ºC and at pH 4.0, respectively (Gomes *et al.*, 2007), while we have obtained at 30ºC and pH 4.0, 6.5% and 8.2%, respectively. Nevertheless, the trend of increased ambiguity when *C. albicans* was exposed to non-optimal growth conditions was maintained.

The incorporation of leucine at CUG codons in *C. albicans* exposed to sorbitol and H_2O_2 was variable, there was no correlation between the increases in

sorbitol concentration and H_2O_2 concentration and leucine misincorporation levels. This observation could indicate that *C. albicans* respond in different ways to different stressors. Previous studies showed that *C. albicans* strains expressing a *Saccharomyces cerevisiae* Leu-tRNA$_{CAG}$ gene that increases leucine incorporation in CUG codons up to 30% displayed extensive morphological variation and production of heterogeneous cell populations that included elongated-ovoid cells, pseudohypha and hypha. These highly ambiguous cells also showed increased cell adhesion, flocculation and production of hydrolases (Miranda *et al.*, 2007), pointing to a possible role for increased CUG ambiguity in pathogenesis, this is consistent with the increased leucine incorporation at CUGs under the stress conditions tested in this study.

Leucine incorporation levels at CUG codons were variable and it is difficult to correlate the differences observed with *C. albicans* ecology and interaction with immune cells. However, our data suggests that variable pH values of host micro-niches, or variation of pH during infection may have an impact on leucine incorporation at levels CUG codons.

The highest level of leucine incorporation at CUG codons were observed in *C. albicans* exposed to macrophages and amphotericin B. These data strongly suggests that the combination of macrophages and amphotericin B has a strong negative effect on the expression of the *C. albicans* SerRS or increases the expression of the LeuRS by several fold. Considering that macrophages have a negative effect on the expression of the SerRS it is likely that this effect is potentiated by amphotericin B. In any case, our data strongly suggests that the regulation of the SerRS and LeuRS play fundamental roles in *C. albicans* biology.

2.6. Conclusion

Previous studies have shown that genetic code ambiguity could play an important role in adaptation and that *C. albicans* is able to cope with high ambiguity levels. In this study, we have developed a reporter system to measure leucine incorporation at CUG codons *in vivo* in an easy way. We observed increased of leucine incorporation at CUG codons in various physiological

conditions. Our data does not support previous data suggesting than CUG ambiguity is regulated through a negative feedback mechanism. Rather, our data show that *C. albicans* is highly adapted to CUG ambiguity and that such ambiguity is an important component of the *C. albicans* response to stress. Future studies should clarify whether such ambiguity is relevant for *C. albicans* pathogenesis.

2.7. References

Ashraf,S.S., Ansari,G., Guenther,R., Sochacka,E., Malkiewicz,A., and Agris,P.F. (1999). The uridine in "U-turn": contributions to tRNA-ribosomal binding. RNA. *5*, 503-511.

Barelle,C.J., Manson,C.L., MacCallum,D.M., Odds,F.C., Gow,N.A., and Brown,A.J. (2004a). GFP as a quantitative reporter of gene regulation in *Candida albicans*. Yeast *21*, 333-340.

Barelle,C.J., Manson,C.L., MacCallum,D.M., Odds,F.C., Gow,N.A., and Brown,A.J. (2004b). GFP as a quantitative reporter of gene regulation in *Candida albicans*. Yeast *21*, 333-340.

Cannon,R.D., Lamping,E., Holmes,A.R., Niimi,K., Tanabe,K., Niimi,M., and Monk,B.C. (2007). Candida albicans drug resistance another way to cope with stress. Microbiology *153*, 3211-3217.

Chauhan,N., Latge,J.P., and Calderone,R. (2006). Signalling and oxidant adaptation in *Candida albicans* and *Aspergillus fumigatus*. Nat. Rev. Microbiol. *4*, 435-444.

Cormack,B.P., Bertram,G., Egerton,M., Gow,N.A., Falkow,S., and Brown,A.J. (1997). Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in *Candida albicans*. Microbiology *143 (Pt 2)*, 303-311.

Davis,D. (2003). Adaptation to environmental pH in *Candida albicans* and its relation to pathogenesis. Curr. Genet. *44*, 1-7.

Dennison,P.M., Ramsdale,M., Manson,C.L., and Brown,A.J. (2005). Gene disruption in *Candida albicans* using a synthetic, codon-optimised Cre-loxP system. Fungal. Genet Biol. *42*, 737-748.

Fernandez-Arenas,E., Cabezon,V., Bermejo,C., Arroyo,J., Nombela,C., ez-Orejas,R., and Gil,C. (2007). Integrated proteomics and genomics strategies bring new insight into *Candida albicans* response upon macrophage interaction. Mol. Cell Proteomics. *6*, 460-478.

Gasch,A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Giege,R. and Lapointe,J. (2009). Transfer RNA Aminoacylation and Modified Nucleosides. In: DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution, ed. Henri GrosjeanAustin, Texas, USA: Landes Bioscience, 475-492.

Giege,R., Sissler,M., and Florentz,C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. *26*, 5017-5035.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206.

Hirasawa,T., Ashitani,K., Yoshikawa,K., Nagahisa,K., Furusawa,C., Katakura,Y., Shimizu,H., and Shioya,S. (2006). Comparison of transcriptional responses to osmotic stresses induced by NaCl and sorbitol additions in *Saccharomyces cerevisiae* using DNA microarray. J. Biosci. Bioeng. *102*, 568-571.

Kumamoto,C.A. (2008). Niche-specific gene expression during C. albicans infection. Curr. Opin. Microbiol.*11*, 325-330.

Lewis,L.E., Bain,J.M., Lowes,C., Gillespie,C., Rudkin,F.M., Gow,N.A., and Erwig,L.P. (2012). Stage specific assessment of *Candida albicans* phagocytosis by macrophages identifies cell wall composition and morphogenesis as key determinants. PLoS. Pathog. *8*, e1002578.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A Genetic Code Alteration Is a Phenotype Diversity Generator in the Human Pathogen *Candida albicans*. PLoS. ONE. *2*, e996.

Noble,S.M. and Johnson,A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. Eukaryot. Cell *4*, 298-309.

Rand,J.D. and Grant,C.M. (2006). The thioredoxin system protects ribosomes against stress-induced aggregation. Mol. Biol. Cell *17*, 387-401.

Ribas de Pouplana,L. and Schimmel,P. (2001). Two Classes of tRNA Synthetases Suggested by Sterically Compatible Dockings on tRNA Acceptor Stem. Cell *104*, 191-193.

Rocha,R., Barbosa Pereira,P.J., Santos,M.A., and Macedo-Ribeiro,S. (2011a). Purification, crystallization and preliminary X-ray diffraction analysis of the seryl-tRNA synthetase from *Candida albicans*. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. *67*, 153-156.

Rocha,R., Pereira,P.J., Santos,M.A., and Macedo-Ribeiro,S. (2011b). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc. Natl. Acad. Sci. U. S. A *108*, 14091-14096.

Santos,M.A., Keith,G., and Tuite,M.F. (1993). Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. EMBO J. *12*, 607-616.

Santos,M.A., Perreau,V.M., and Tuite,M.F. (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. EMBO J. *15*, 5060-5068.

Soma,A., Kumagai,R., Nishikawa,K., and Himeno,H. (1996). The anticodon loop is a major identity determinant of *Saccharomyces cerevisiae* tRNA(Leu). J. Mol. Biol. *263*, 707-714.

Suzuki,T., Ueda,T., and Watanabe,K. (1997). The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122-1134.

Walther,A. and Wendland,J. (2003). An improved transformation protocol for the human fungal pathogen *Candida albicans*. Curr. Genet *42*, 339-343.

3. Characterization of ambiguous *Candida albicans* strains

3.1. Abstract

The ascomycete *Candida albicans* is a normal resident of the gastrointestinal tract of humans and other warm-blooded animals. It is a successful commensal and a pathogen that occurs in a broad range of body sites. It also has high capacity to survive and proliferate in environments with drastic changes in oxygen, carbon dioxide, pH, osmolarity, nutrients and temperature (Gasch, 2007). Surprisingly, it has a unique translation system as it translates leucine CUG codons as serine (Santos and Tuite, 1995). Under standard growth conditions 3% of leucine and 97% of serine are incorporated at CUGs, but these values are flexible as leucine misincorporation fluctuates between 0.6% and 5% in response to environmental stress (Gomes *et al.*, 2007). Misincorporation of leucine can be artificially increased up to 100%, with small visible effects on fitness at ambiguity levels up to 28% (Gomes *et al.*, 2007). In order to determine whether such genetic code flexibility exists at other codons sites, we have produced *C. albicans* strains that ambiguously translate CTA , CTC, CTT (Leucine), ATC (Isoleucine), GCC (Alanine), GGA (Glycine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) codons as serine. These strains show decrease in fitness but can tolerate ambiguity at other codons. Some strains showed a small decrease in growth rate but an unanticipated increase in the levels of protein synthesis. These interesting results led us to characterize these ambiguous strains using phenotypic profiling in variable growth conditions, including temperature, pH, nutritional cues and antifungal drugs. Other tests included the analysis of nuclear DNA content and detection of insoluble protein aggregates. The majority of the strains showed growth advantages on different carbon sources, osmotic stress and oxidative stress, suggesting that ambiguity plays an important role in adaptation to new environmental conditions.

3.2. Introduction

In 1968, Crick defined the genetic code as “a non-overlapping triplet code. Most, but not all, of the 64 triplets stand for one or another of the 20 amino acids

and, in most cases, each amino acid is represented by more than one codon" (Crick, 1968). This definition of the genetic code represents a modest subset of the overall complexity of the translation machinery, namely mRNA, tRNAs, amino acids, aminoacyl-tRNA synthetases, ribosome and termination factors. It also freezes the code in a non-evolvable status.

The discovery in 1979 that human mitochondria translate the UGA stop codon as tryptophan and the AUA isoleucine codon as methionine (Barrell *et al.*, 1979) showed that the code evolves. However, this change in the code was viewed as a special case due to the small size of the genome of human mitochondria (Ohama *et al.*, 2008). This idea was put into question in 1985 with the discovery that *Mycoplasma capricolum* translates the UGA stop codon as tryptophan (Yamao *et al.*, 1985). Since then variation to the standard genetic code has been reported in numerous organisms, totalizing 16 codon changes in mitochondrial codes and 10 codons in bacterial and nuclear codes (Knight *et al.*, 2001;Di, 2005). Although these alterations occur in different organisms, several of them affect the same codons, namely the stop codons UAA, UAG and UGA. In numerous green algae, ciliates and *Diplomonads* the nuclear codons UAA and UAG encodes glutamine. The codon UGA encodes tryptophan in *Mycoplasma* species, *Spiroplasma citri, Bacillus* and some ciliates; cysteine in *Euplotes*; and in *Pseudomicrothorax dubius* and *Nyctotherus ovalis* an unidentified amino acid (Soll and RajBhandary, 2006). Furthermore, some codons are used at very low level or are absent in the genome and are not assigned to any amino acid, namely the arginine AGA codon and the isoleucine AUA codon in *Micrococcus* species (Kano *et al.*, 1993), and arginine CGG codon in *Mycoplasma capricolum* (Oba *et al.*, 1991).

In mitochondria, genetic code alterations are widespread and a large number of phyla, with the exception of green plants, use the stop codon UGA to incorporate tryptophan into proteins. The other stop codon UAA is assigned to tyrosine in several *Platyhelminth*, while UAG is assigned to alanine and leucine in fungi and several plants. The lysine AAA codon is assigned to arginine in *Platyhelminth* and *Echinoderm*, and the arginine codons AGA and AGG are used as termination codons in various vertebrates and as serine codons in various

animal mitochondria. Some organisms use AGA for glycine while others use it for serine. The serine UCA codon is also a stop codon in the green alga *Scenedesmus obliques* (Knight *et al.*, 2001;Soll and RajBhandary, 2006).

In recent years, two naturally occurring amino acids were added to the canonical 20 amino acids, namely selenocysteine, which is designated as the 21st amino acid. It is found in various species from the 3 kingdoms of life and is incorporated into selenoproteins by reprograming the UGA codon (Ambrogelly *et al.*, 2007). The 22nd amino acid is pyrrolysine, which is incorporated into proteins in response to the UAG codon in *Methanosarcinaceae* (Srinivasan *et al.*, 2002;Soll and RajBhandary, 2006).

These widespread genetic code alterations show that the code is neither universal nor frozen. Indeed, recent studies have demonstrated that bacterial, fungal and mammalian cells are highly tolerant to incorporation of artificial amino acids, using orthogonal tRNA-synthetase pairs (Cropp and Schultz, 2004). For example, Wang and coworkers inserted in *Escherichia coli* an orthogonal tyrosyl-tRNA synthetase (TyrRS) from *Methanococcus jannaschii* and a mutant tyrosine amber suppressor tRNA (Wang *et al.*, 2001). This pair incorporates the synthetic amino acid *O*-methyl-L-tyrosine into proteins in response to an amber nonsense codon (UAG). Since the orthogonality prevents the new tRNA-synthetase (aaRS) from aminoacylating cellular tRNAs and the new tRNA from being recognized by the host tRNA-synthetases, the system is highly specific (Wang *et al.*, 2001). Anderson and co-workers were also able to engineer an orthogonal aaRS-tRNA pair derived from archaeal tRNALys which efficiently and selectively incorporates a non-canonical amino acid into proteins in response to the quadruplet codon AGGA (Anderson *et al.*, 2004). Genetic code engineering has also been achieved using an editing defective aaRS. Deletion of the editing domain of an isoleucyl-tRNA synthetase (IleRS) resulted in ambiguous translation of isoleucyl codons with non-canonical amino acids (Pezo *et al.*, 2004). Finally, tRNAs bearing mutations in the anticodon that do not affect aminoacylation have been used to introduce various amino acids at non-cognate codons. For example serine has been incorporated at 19 different codons in chick embryos and human cells (Geslain *et al.*, 2010).

Interestingly, *C. albicans* uses a mutant tRNA to incorporate serine at CUG codons (Santos and Tuite, 1995). This mutant $tRNA_{CAG}^{Ser}$ is a typical serine tRNA and is recognized by the seryl-synthetase (serRS), but has 3 mutations in the anticodon that allow it to read the leucine CUG codon as serine. Interestingly, the anticodon mutation (A_{35} and m'G_{37}) are recognized by the leucyl-tRNA synthetase (LeuRS) and the $tRNA_{CAG}^{Ser}$ is also charged with leucine (Soma *et al.*, 1996), originating two aminoacyl-tRNAs, namely the Ser-$tRNA_{CAG}^{Ser}$ and Leu-$tRNA_{CAG}^{Ser}$. Since Leu-$tRNA_{CAG}^{Ser}$ is not edited by the LeuRS nor discriminated by the translation elongation factor 1 (eEF1A) (Santos *et al.*, 1997;Santos *et al.*, 2011), it participates in protein synthesis, creating ambiguity at CUG codons (Gomes *et al.*, 2007).

Under standard growth conditions 3% of leucine and 97% of serine are incorporated at CUGs (Gomes *et al.*, 2007). However, these values are flexible and leucine misincorporation fluctuates between 0.6% and 5% in response to environmental stress. Remarkably, misincorporation of leucine can be artificially increased up to 28% with small visible effects on fitness (Gomes *et al.*, 2007). In order to determine if such genetic code flexibility exists at other codons sites, we have produced *C. albicans* strains that incorporate serine at several non-cognate sites. For this, we have altered the anticodon of a serine tRNA producing chimeric tRNAs similar to the natural $tRNA_{CAG}^{Ser}$ that incorporates serine at CUG leucine sites. We have produced serine tRNAs with anticodons that read CTA , CTC, CTT (Leucine), ATC (Isoleucine), GCC (Alanine), GGA (Glycine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) codons. Most strains exhibited little decrease in fitness indicating that *C. albicans* tolerates ambiguity at non-CUG codons. Some strains did have slow growth rate, but surprisingly an unanticipated increase in protein synthesis rate was observed. Phenotypic profiling testing growth in variable conditions, namely temperature, pH, antifungal drugs and nutritional cues, showed that most strains had growth advantages in presence of osmotic stress, oxidative stress and different carbon sources. These growth advantages suggest that ambiguity plays an important role in adaption to new environmental conditions. Other phenotypic differences between strains were observed, including protein aggregation in the strain misincorporating serine at leucine CTC codons and

increased DNA content in strains misincorporating serine at leucine CTA and CTC codons.

3.3. Material and Methods

3.3.1. Strain maintenance

The *C. albicans* strain SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434}) (Noble and Johnson, 2005) was a gift from Alexander D. Johnson (University of California-San Francisco). SN148 was grown in YPD (1% yeast extract, 2% peptone, 2% dextrose) at 30ºC. Transformed strains were grown in synthetic minimal medium (0.67% yeast nitrogen base, 2% glucose, 0.2% Drop-out mix with all the essential amino acids). *E. coli* JM109 (*recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, Δ(*lac-proAB*)/F' [*traD*36, *proAB*$^{+}$, *lacI*^{q}, *lacZ*ΔM15]) was grown in LB (1% peptone from casein, 0.5% yeast extract, 1% sodium chloride) at 37ºC, when necessary LB was supplemented with ampicillin (75 mg/ml).

3.3.2. Plasmid purification and construction

The plasmids used in this study were purified using GeneJet^{TM} Plasmid Miniprep Kit (Fermentas) according to the manufacturer's instructions. Plasmid pUA552 was constructed by PCR amplification of tDNAser (UGA) from *S.cerevisiae*, with oligo oUA1671 5'-AAAGGTACCGAAGGAGGTGCAAGGGAAAAG-3' and oUA1672 5'-TTTGGGCCCTCCGTGCATAACGAATGACTC-3' and inserted between *KpnI* and *ApaI* restriction enzyme sites of PMG2287 (gift from Prof. Judith Berman). The tRNA was mutated using the QuikChange Site-Directed Mutagenesis kit (Stratagene) with oligos oUA1766 5'-TAAGGCGACAGACGTGAAATCTGTTGGGCTC-3' and oUA1767 5'-

GAGCCCAACAGATTTCACGTCTGTCGCCTTA-3' with the purpose of exchanging T33 for G33 to decrease serylation efficiency of the tRNA (Santos *et al.*, 1996). Then plasmid pUA552 (tRNA anticodon) was mutated using the oligonucleotides described in Table 3-1 to produce new plasmids. These plasmids were linearized with *StuI* and integrated into the *C. albicans* genome.

Table 3-1: List of plasmids and primer sequences used to mutate the $tRNA^{Ser}$ anticodon. The new anticodons obtained by mutagenesis and corresponding codons and amino acids are also indicated

Plasmid Name	Primer Name	Primer Sequence (5'->3')	Anticodon (5'->3')	Codon (5'->3')	Amino acid
pUA533	oUA1719	TAAGGCGACAGACGTAGAATCTGTTGGGCTC	UAG	CTA	Leucine
	oUA1720	GAGCCCAACAGATTCTACGTCTGTCGCCTTA			
pUA534	oUA1721	TAAGGCGACAGACGGAGAATCTGTTGGGCTC	GAG	CTC	Leucine
	oUA1722	GAGCCCAACAGATTCTCCGTCTGTCGCCTTA			
pUA535	oUA1723	TAAGGCGACAGACGAAGAATCTGTTGGGCTC	AAG	CTT	Leucine
	oUA1724	GAGCCCAACAGATTCTTCGTCTGTCGCCTTA			
pUA536	oUA1725	TAAGGCGACAGACGGATAATCTGTTGGGCTC	GAU	ATC	Isoleucine
	oUA1726	GAGCCCAACAGATTATCCGTCTGTCGCCTTA			
pUA537	oUA1727	TAAGGCGACAGACGGGCAATCTGTTGGGCTC	GGC	GCC	Alanine
	oUA1728	GAGCCCAACAGATTGCCCGTCTGTCGCCTTA			
pUA540	oUA1739	TAAGGCGACAGACGTCCAATCTGTTGGGCTC	UCC	GGA	Glycine
	oUA1740	GAGCCCAACAGATTGGACGTCTGTCGCCTTA			
pUA542	oUA1747	TAAGGCGACAGACGCTTAATCTGTTGGGCTC	CUU	AAG	Lysine
	oUA1748	GAGCCCAACAGATTAAGCGTCTGTCGCCTTA			
pUA544	oUA1751	TAAGGCGACAGACGGGTAATCTGTTGGGCTC	GGU	ACC	Threonine
	oUA1752	GAGCCCAACAGATTACCCGTCTGTCGCCTTA			
pUA546	oUA1755	TAAGGCGACAGACGGTAAATCTGTTGGGCTC	GUA	TAC	Tyrosine
	oUA1756	GAGCCCAACAGATTTACCGTCTGTCGCCTTA			

3.3.3. *C. albicans* transformation

Transformation of *C. albicans* was carried out using an improved lithium acetate method (Walther and Wendland, 2003) with small modifications. Briefly, overnight cultures were inoculated in fresh medium to an $O.D_{600nm}$ of 0.3 and let

grow at 30ºC, 200rpm shaking until an $O.D_{600nm}$ of 1-1.2. Cell cultures were then transferred to 50ml falcon tubes and centrifuged at 3220 rcf (4000rpm) for 5min, supernatants were discarded and pellets were resuspended in 1.5ml of LiAc-solution (0.1M LiAc, TE buffer 1x). 200 µl of suspension were transferred to 1.5mL eppendorf tubes and the transformation reagents were added in the following order: 600µL 50% (w/w) PEG LiAc-solution, 50µL single-stranded carrier DNA (2mg/mL) previously denatured and 50µL of an aqueous solution of the plasmid of interest (containing 1– 5µg of plasmid). Tubes were vortexed and incubated at 30ºC for 4 hours following incubation at 44ºC for 15min and then placed on ice for 2min. Cells were harvested at 4000rpm for 5min and resuspended gently in 200µl of appropriate SC selective medium. Each 100µl of suspension was plated onto appropriate SC selection medium plates and incubated at 30°C for 3-4 days.

3.3.4. Plasmid integration

Correct integration at the *RPS10* locus was confirmed by PCR using the primers oUA1554 5'- CGTATTCACTTAATCCCACAC-3', oUA1555 5'-CCAATTGGTGATGGTCC-3'. Single copy integration was selected using the primers oUA1556 5'-GGTATAGAAATGCTGGTTGG-3' and oUA1557 5'-CCAATTGGTGATGGTCC-3', which only obtain a PCR amplification product for tandem integrated plasmids (Barelle *et al.*, 2004b).

3.3.5. Growth rate

C. albicans transformed strains were grown until late stationary phase at 30ºC with an agitation of 180 rpm. Next, strains were inoculated to an initial $O.D_{600nm}$ of 0.02 with 20 ml of minimal medium lacking uridine in 100 ml Erlenmeyer flasks. Cultures were grown at 30ºC at 180 rpm of agitation and culture $O.D_{600nm}$ was measured at several time points. A growth curve was determined by plotting exponential growth phase values in a logarithmic scale. These values were used for a linear regression and the slope of the obtained

equation corresponds to the growth rate value. At least three independent experiments using three different clones were performed for each strain.

3.3.6. Total RNA extraction

About 25 OD_{600nm} units (volume of culture X OD_{600nm}) of exponentially growing cells (OD_{600nm} = 0.5~0.8) were harvested by centrifugation at 4000 rpm for 4 minutes at room temperature. After removal of the supernatant, tubes were immediately immersed in liquid nitrogen for 6 minutes, frozen and stored at -80ºC. Total RNA was isolated from yeast using an hot phenol based protocol (Schmitt *et al.*, 1990a). Mainly, frozen pellets were resuspended in 500µl acid phenol chlorophorm (Sigma, 5:1, pH 4.7) at 65ºC. An equal volume of TES-buffer (10 mM Tris pH 7.5, 10 mM EDTA and 0.5% sodium dodecyl sulfate) was added and each tube was vortexed at high speed for 20 seconds to resuspend the pellets. Tubes were then incubated for 1 hour in a water bath at 65ºC with 20 seconds vortexing every 10 minutes. Tubes content were transferred to 1.5 ml tubes and centrifuged for 20 minutes at 14000 rpm at 4ºC. The water-phase was added to a new tube, filled with 600 µl Acid Phenol Chloroform (Sigma, 5:1, pH 4.7), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm, at 4ºC. For a second extraction the water-phase from this step was added to new Eppendorf tubes filled with 500 µl Chloroform:Isoamyl-alcohol (Sigma, 25:1) which were vortexed for 20 seconds and centrifuged for 10 minutes, at 14000 rpm, at 4ºC. Again, the water-phase was transferred to a new Eppendorf tube filled with 400 µl Chloroform:Isoamyl-alcohol (Sigma, 25:1), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm, at 4ºC. Finally the water-phase was added to new Eppendorf tubes with 35 µl sodium acetate (3 M, pH 5.2), 800 µl ethanol (100 %, kept at – 20ºC) and incubated at – 20ºC for at least 1 hour. Tubes were then centrifuged for 5 minutes at room temperature, 14000 rpm, supernatants were removed carefully with a pipette tip, avoiding touching the RNA-pellet. Pellets were then washed with 500 µl ethanol (80 %, -20ºC) and centrifuged for 3 minutes at room temperature, 14000 rpm. Ethanol was removed

and the RNA pellets were air dried for 1 minute. RNA pellets were dissolved in sterile (mQ) water to a concentration of 10 μg / μl. Samples were stored at -80ºC.

3.3.7. Northern blot analysis

Briefly, 50 μg of total RNA were resolved at room temperature in 15% polyacrylamide (40% Acril:Bis) gels containing 8 M urea, buffered with 1X TBE pH 8.0, electrophoresed at 500 V for 16 hours. The section of the gels containing the tRNAs was transferred to a nitrocellulose membrane (Hybond N, Amershan) by a Semi-Dry Trans Bloting system (Bio-Rad). Probes were prepared by phosphorylating 10 pmol of dephosphorylated oligonucleotide (Table 3-2) with 4 μl of γ-^{32}P-ATP (5000Ci/mmol) (Perkin Elmer) in 1X T4 Kinase buffer, 10 mM spermidine and 16 units of T4 Kinase (Takara). This reaction was carried out by incubating the reactions for 1 h at 37ºC. Probes were purified with 100 μl phenol:chlorophorm:isoamyl alcohol (PCIA) and hybridized in a hybridization solution (6X SSPE, 5X Denhart's solution, 0.05% sodium dodecyl sulfate), overnight at 52ºC (Y.Espanõl, unpublished). Hybond-N membranes were washed with washing solution (2X SSPE, 0.5% sodium dodecyl sulfate), wrapped in a plastic bag and exposed for 24 hours to a K-screen and scanned using Molecular Imager FX (Bio-Rad) with adequate settings.

Table 3-2: List of oligonucleotide used as probes for the northern blot analysis.

Oligo ID	Detect	Sequence 5'-3'
oUA1699	$tRNA_{UGU}^{Thr}$	ACGCTCTACCACTAAGCTAA
oUA1757	$tRNA_{UGA}^{Ser}$	TTAACCACTCGGCCATAGT

3.3.8. Protein synthesis quantification

Protein synthesis quantification was performed as described formerly with few alterations (Alamgir *et al.*, 2008). Briefly, 2X10^8 cells from mid exponential growing phase were collected and washed three times with 30ºC minimal medium

lacking Met and Uri. Cells were then resuspended in 2 ml of pre-warmed minimal medium lacking Met and Uri and incubated for 20 min at 30ºC with agitation. 2 µl of cold L-[^{35}S]-Methionine (Perkin Elmer, 1175 Ci/mmol, 10.5 mCi/ml) were added and the mixture was incubated for 8 min at 30ºC with agitation. Amino acid incorporation was stopped by addition of 60µl of cycloheximide (20mg/ml) and incubation on ice. Cells were washed three times with cold water and were frozen at -80ºC. Protein was then extracted by resuspending cell pellets in 200µl Lysis buffer [50mM potassium phosphate buffer pH 7, 1 mM EDTA, 5% (vol/vol) glycerol, 1mM phenylmethylsulfonyl fluoride, and complete mini protease inhibitor cocktail (Roche)] and 120µl of glass beads. Cells were disrupted using a Precellys disrupter (5 cycles of 10 sec at 5000rpm and 1min on ice between cycles) and centrifuged at 3000xg for 10min. 30µl of supernatant were applied onto 1 cm^2 square paper microfiber filters (GF/C, Whatman). Radioactivity counting was performed using a scintillation counter (Beckman). Protein extracts were quantified using the BCA protein quantification Kit (Pierce). Counting was normalized against the total protein for each sample and compared to control.

3.3.9. Flow cytometry

Flow cytometry analysis was performed as follows. Cells were grown overnight in YPD at 30ºC, then concentrated to a density of 1x10^7 cells/ml and centrifuged at 10.000 rpm for 3 min. Cells were then washed with 1ml of PBS, fixed in 1ml of cold (-20ºC) ethanol 70% and incubated overnight at 4ºC. Fixed cells were washed and resuspended in 700 µl of RNase A (Sigma) (1mg/ml) and incubated at 50ºC for 1 h, 50 µl of proteinase K (20mg/ml) were then added and the mixture was incubated for 1 h at 50ºC. Next, 50 µl of propidium Iodide (1mg/ml) were added and samples were incubated overnight at 4ºC in the dark. Finally, samples were analyzed using a flow cytometer Beckman Coulter with the proper filters and settings. Data analysis was performed using FlowJo software (www.flowjo.com).

3.3.10. Phenotypic assays

Phenotypic assays were performed as previously described (Homann *et al.*, 2009) with minor alterations. Briefly, $1X10^8$ cells from mid exponential phase cultures were collected and resuspended in 1ml of water. Six 10X dilutions were transferred to 96-well micro plates. This format allowed for cell plating using a liquid handling station (Caliper LifeSciences). The assay plates (agar) were incubated for 5 days and colonies were photographed using a dissecting microscope equipped with AxioCam HRc camera and Axio Vision Software from Zeiss. Images were analyzed using Image J and colonies size was determined. Growth scores (GS) of each strain were calculated by dividing the size of the colonies in the assay plate by the size of the colonies in the control plates, for the corresponding dilutions (Equation 3.1 and 3.2). From these growth scores of mistranslating strains (GS_m) the growth score of control strain (GS_c) was subtracted, producing relative growth score (RGS) for each mistranslating strain in each assay. Assay plates conditions are described in Table 3-3.

Table 3-3: List of stressor compound with respective concentration and conditions used.

Assay	Stress compound	Concentration	Base medium	Growth temperature
Control			YEPD/ MM-Uri	30ºC
Temperature			MM-Uri	25ºC 37ºC 42ºC
Elevated cation concentration	Calcium chloride Sodium chloride	300 mM 1.3 M	MM-Uri	30ºC
Carbon source	Carbon source absent Galactose Glycerol Ethanol	 2% (w/v) 3% (w/v) 2% (w/v)	MM-Uri without glucose	30ºC
Protein denaturation	Guanidine HCl Urea	5 mM 25 mM	MM-Uri	30ºC
pH value	pH 5.0 pH 8.6	pH 5.0 pH 8.6	YEPD + buffered glycine	30ºC
Stress	Calcofluor White Sorbitol Caffeíne EDTA Hydrogen Peroxyde $CuSO_4$	20 µM 1.5 M 15 mM 0.75 mM 6.0 mM 13 mM	MM-Uri MM-Uri MM-Uri YEPD/ MM-Uri YEPD/ MM-Uri YEPD	30ºC
Antifungal resitance	Fluconazole	0.5 µg/ml	MM-Uri	30ºC

- GS_m = Growth score of mistranslating strain;
- GS_c = Growth score of pUA552 control strain;
- A_{id} = Measured area for isolate “i” and spot dilution “d”;
- n_{id} = the set of all isolates in a given dilution for the strain under study;

$$GS_m = \frac{1}{n_{id}} \sum_{n \in id} \left(\frac{[A_{id}]stress}{[A_{id}]nostress} \right)$$

Equation 3.1

$$GS_c = \frac{1}{n_{id}} \sum_{n \in id} \left(\frac{[A_{id}]stress}{[A_{id}]nostress} \right)$$

Equation 3.2

3.3.11. Insoluble protein quantification

Quantification of insoluble protein was carried as described by (Rand and Grant, 2006) with minor modifications. Cells were grown to exponential phase in MM-Uri. Equal number of cells were harvested by centrifugation, washed and transferred to 2 ml microtubes and frozen at -80ºC for 30 min. Tubes were removed from the freezer and 300µl of lysis buffer [50 mM potassium phosphate buffer, pH 7, 1 mM EDTA, 5% (vol/vol) glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF), and Complete Mini protease inhibitor cocktail (Roche)] plus lyticase (20 mg/ml) were added to each tube. Reaction mixtures were incubated for 30 min at 37ºC. Cells were disrupted with 100µl of glass beads using a Precellys 24 (bertin technologies) 3X at 5000 rpm for 10 sec, followed by 2 min of incubation on ice. Mixtures were then centrifuged at 3000g for 15 min at 4ºC, to remove intact cells, and supernatants were transferred to new tubes. Another centrifugation was carried out at 15000xg for 20 min to isolate membranes and

protein aggregates. Next, membrane proteins were removed by washing pellets twice with 320µl of lysis buffer complemented with 80 µl of 10% Triton-X100, centrifuging at 15000xg for 20 min each time. Finally aggregated proteins were resuspended in 50 µl of lyses buffer and 10 µl of 6 X sample buffer (30% glycerol, 10% sodium dodecyl sulfate, 0.6 M DTT, and 0.012 bromophenol blue in 0.5 M Tris-Cl / 0.4 % SDS, pH 6.8) and boiled for 5 min at 95ºC. Aliquots of 10 µl of each sample were fractionated on 15% acrylamide gels. Following electrophoresis, gels were stained by submersion on a coomassie blue solution (0.25 % Brilliant Blue R, 50 % methanol and 10 % acetic acid) for 2 hours with slow agitation. Destaining was carried out using 25% methanol and 5% acetic acid solutions, overnight with slow agitation. Gels were then washed with distilled water and scanned using the ODYSSEY Infrared Imaging System (Li-Cor Biosciences) with the adequate definitions. Gel images obtained were analyzed with ImageJ software and gel bands intensity was quantified.

3.4. Results

3.4.1. Selection of codons for mistranslation studies

C. albicans strains were engineered to misincorporate serine at various non-cognate codons, using mutant serine tRNAs (Figure 3-1). For this, the *S. cerevisiae* $tRNA_{UGA}^{Ser}$ gene was mutated in its anticodon to produce anticodons complementary to various non-cognate codons. Since the seryl-tRNA synthetase (SerRS) does not recognize the $tRNA^{ser}$ anticodon (it binds to the acceptor stem, D-arm and extra stem/loop), the alterations in the anticodon of the $tRNA_{UGA}^{Ser}$ did not alter the charging specificity (Lenhard *et al.*, 1999). In other words, the mutated tRNAs (chimeric tRNAs) were aminoacylated by SerRS leading to serine incorporation at positions encoded by codons complementary to the engineered anticodons. These chimeric $tRNAs^{Ser}$ competed with the native tRNAs resulting in statistical incorporation of serine in the selected codons sites. In other words, serine misincorporation produced statistical proteins.

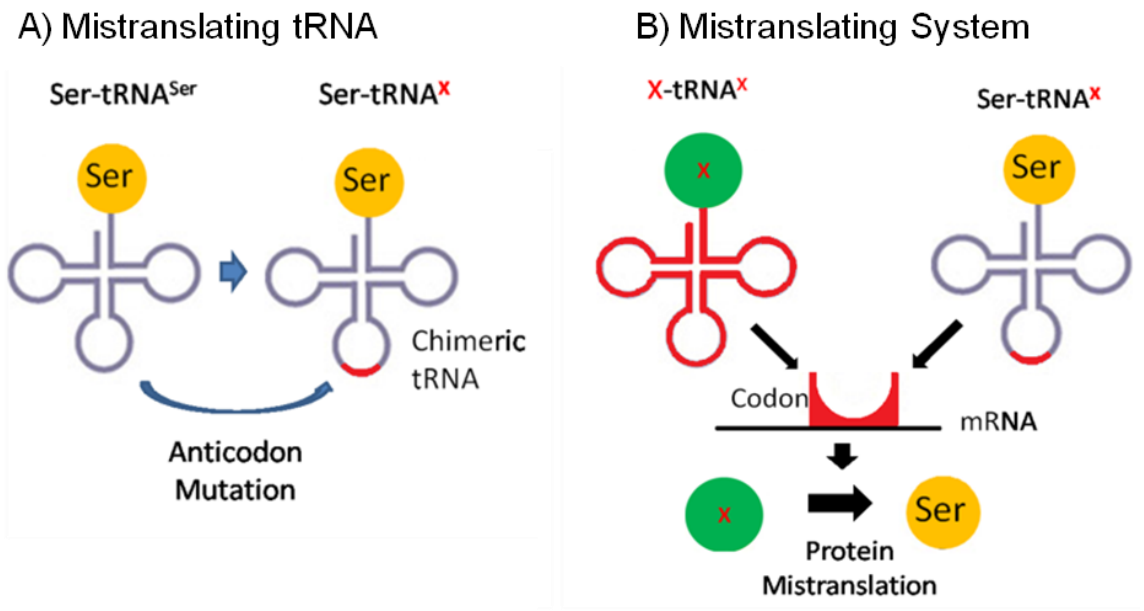


Figure 3-1: Novel mistranslation scheme, A) Mistranslating tRNA, the *S. cerevisiae* $tRNA_{UGA}^{Ser}$ anticodon was mutated to produce a chimeric $tRNA^{Ser}$ that inserts serine at non-cognate codons. B) Mistranslating system. The serylated chimeric Ser- $tRNA^{X}$ competes with the cognate X-$tRNA^{X}$ for the targeted codon, serine misincorporation results in production of mutant (statistical) proteins.

The $tRNA^{Ser}$ anticodon mutations were selected to ensure that serine was misincorporated at sites encoded by codons whose usage levels corresponds approximately to 10 % of the targeted amino acid. Amino acids chemical properties were also taken into consideration (Table 3-4). The engineered *C. albicans* strains misincorporated serine at CTA , CTC, CTT (Leucine), ATC (Isoleucine), GCC (Alanine), GGA (Glycine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) codons. Leucine and isoleucine were chosen on the basis of the differences in molecular volume, molecular weight and hydropathy, being the most deleterious mistranslations. Threonine, glycine and alanine were selected because they have similar molecular volume to serine, producing low toxicity effects. Tyrosine was selected since it is an aromatic amino acid. Finally, lysine was chosen from the polar basic amino acid group.

Table 3-4: Amino acids characteristics and codon usage. Table adapted from Haig and Hurst 1991 (Haig and Hurst, 1991).

Amino acid	Codon (5'->3')	codon usage (frequency per 1000)	Hydropathy	Molecular volume (Å^3)	Molecular Weight (D)
Alanine (Ala)	GCC	11.7	1.8	88.6	89
Glycine (Gly)	GGA	13.7	-0.4	60.1	75
Isoleucine (Ile)	ATC	13.5	4.5	166.7	131
Leucine (Leu)	CTC	2.6	3.8	166.7	131
	CTA	4.4			
	CTT	10.2			
Lysine (Lys)	AAG	18.3	-3.9	168.6	146
Serine (Ser)	TCA	26.4	-0.8	89	105
Threonine (Thr)	ACC	13.5	-0.7	116.1	119
Tyrosine (Tyr)	TAC	10.4	-1.3	193.6	181

3.4.2. Phenotypic and cellular consequences of tRNA mistranslation

With the purpose of evaluating the impact of the chimeric tRNAs on *C. albicans* fitness, several parameters were analysed. Yeast fitness assays included the determination of transformation efficiencies and growth rates. To elucidate the extension of the impact of serine misincorporation in *C. albicans* physiology, the performance of the engineered strains was evaluated in different growth conditions including temperature, pH, nutritional cues and antifungal drugs. Finally, DNA content variation, protein synthesis and protein aggregation were also tested.

There was a strong reduction of the transformation efficiency of *C. albicans* with plasmid carrying chimeric tRNAs (Figure 3-2) relative to the control plasmid, indicating that mistranslation is toxic to *C. albicans*, as one would expect. The plasmids that displayed lower transformation efficiency encoded tRNAs that misincorporated serine at leucine, isoleucine and threonine codons. Surprisingly, the plasmid carrying the wild type (WT) $tRNA^{ser}$ (pUA552) also showed lower transformation efficiency relative to the empty plasmid (PMG2287), suggesting that increasing the expression level of the $tRNA_{UGA}^{Ser}$ is toxic.

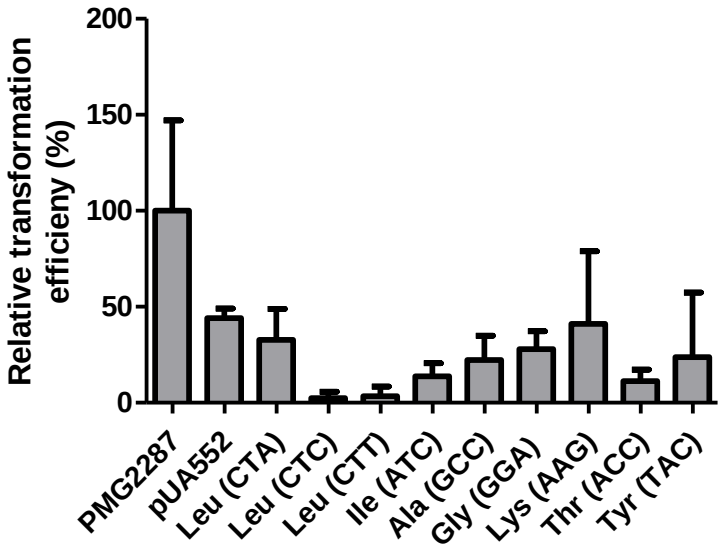


Figure 3-2: Transformation efficiency. Transformation efficiency was determined using at least 3 independent transformations. Values obtained were normalized relative to the control PM2287 strain (%). Data are mean + s.d., and shows sharp reduction in the transformation efficiency of all misreading tRNA genes, relative to the control.

To determine if the different mutant tRNAs were expressed, a northern blot analysis was carried out. The only tRNA that was not detected was the one that mistranslated leucine CTC codons (Figure 3-3). The respective strain showed the lowest transformation efficiency (Figure 3-2) and growth rate (Figure 3-4), suggesting that this chimeric tRNA is highly toxic and its expression is strongly downregulated.

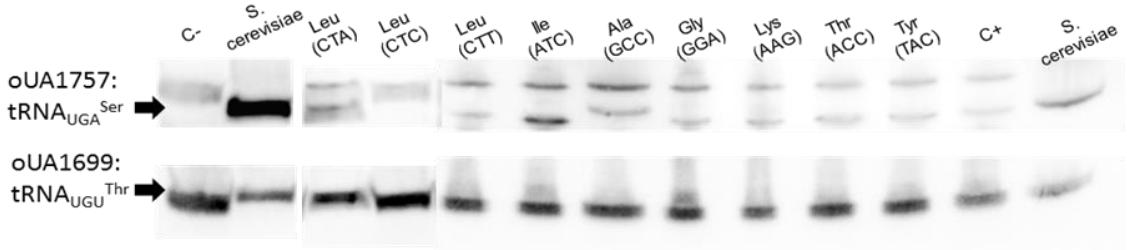


Figure 3-3: Northern blot analysis of the mutant tRNAs. 50 µg of total RNA were resolved at room temperature on 15% polyacrylamide (40% Acril:Bis) gels containing 8 M urea. Probe oUA1757 labeled with ɣ-^{32}P-ATP was used to detect tRNA$_{UGA}^{Ser}$, probe oUA1699 labeled with ɣ-^{32}P-ATP was used to detect the internal control tRNA$_{UGU}^{Thr}$.

To evaluate the impact of mistranslation on *C. albicans* fitness, growth rate was determined. There was a strong negative effect on growth of all strains, however serine misincorporation at chemically distinct amino acids (threonine and

leucine) sites had the highest impact on growth rate. Indeed, strains Leu (CTC) and Thr (ACC) had approximately 50% and 60% decrease in growth rate relative to the control pUA552, respectively. Furthermore, the other strains that revealed a significant decrease in growth rate were Leu (CTA), Leu (CTT), Ile (ATC), Ala (GCC) and Gly (GGA) with 39%, 25%, 29%, 46% and 34% decrease, respectively. These results are in agreement with previous studies where yeast *S. cerevisiae* growth rate was also detrimentally affected by codon ambiguity (Santos *et al.*, 1996).

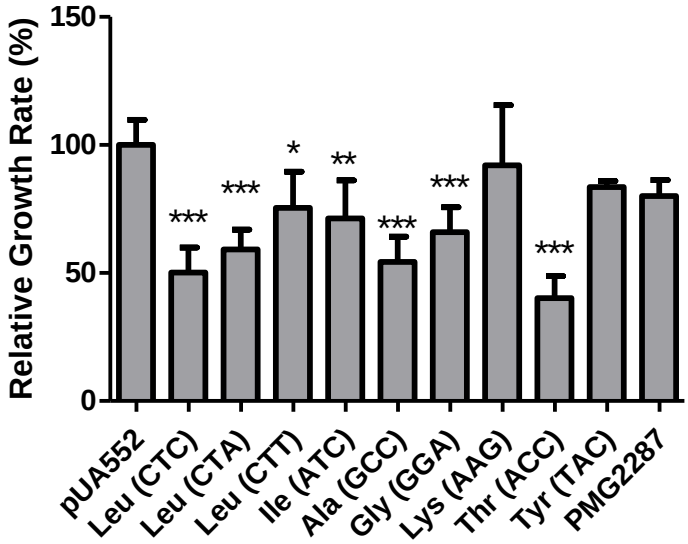


Figure 3-4: Growth rate of mistranslating *C. albicans* strains. The data shows decreased growth rate in all strains, however strains misincorporating serine at chemically distinct sites show sharper negative effects. Data represents growth rate mean + s.d. of triplicates of 3 different clones, normalized to the pUA552 control. Data statistical analysis one-way ANOVA was performed followed by a Dunnet test with CI 95% relative to pUA552 (***p<0.001, **p<0.01, *p<0.05).

When cells are exposed to stress, namely heat shock, osmotic and oxidative stress, protein synthesis decreases. This minimizes production of erroneous proteins and helps reprograming gene expression to adjust it to the demands of the new growth conditions (Gasch *et al.*, 2000;Gasch and Werner-Washburne, 2002;Gasch, 2007) .

Moreover, mistranslation induces proteotoxic stress and a recent study, using an inducible mistranslating *S. cerevisiae* strain, showed decreased protein synthesis rate upon induction of mistranslation (Paredes *et al.*, 2012). In order to

clarify if the different types of mistranslation engineered in this study could affect protein synthesis rate in the already ambiguous *C. albicans*, we quantified it by determining incorporation of [^{35}S]-methionine into proteins *in vivo*. We pulse-labelled proteins by adding 1 µCi of [^{35}S]-methionine to 10^7 of actively growing cells. Protein synthesis was stopped by addition of cycloheximide and crushed ice. Remarkably, there was a decrease in protein synthesis rate in all strains (Figure 3-5). The strongest effect occurred in strains misincorporating serine at leucine (CTC, CTT), alanine (GCC) and, lysine (AAG) codons. In most cases, growth and protein synthesis rates were correlated, except in the strain mistranslating threonine codons.

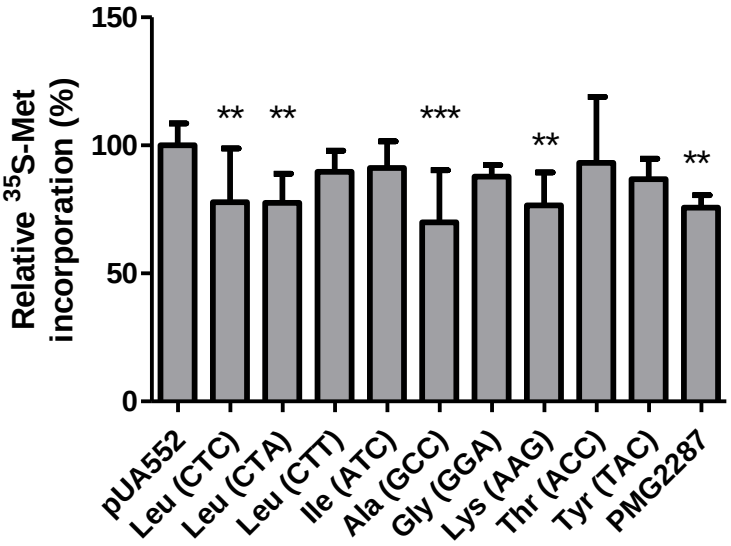


Figure 3-5: Effect of mistranslation on protein synthesis rate. *C. albicans* cells were collected in exponential growth phase and incubated for 20 min in media lacking methionine. Cells were labelled with [^{35}S]-Met for 8 min and protein synthesis was then stopped by adding cycloheximide. Data represents mean incorporation of [^{35}S]-Met on new synthesized protein + s.d. of triplicates of 3 different clones, normalized with pUA552. Statistical analysis was carried out using one-way ANOVA followed by a Dunnet test with CI 95% relative to pUA552 (***p<0.001, **p<0.01, *p<0.05).

Several studies suggested that mistranslation induces protein aggregation. For example, a mouse mutation in the editing domain of the alanyl-tRNA synthetase disrupts the enzyme aminoacylation proofreading activity, generating misacylated Ser-tRNA^{Ala} and widespread translation errors. This leads to the production of misfolded proteins and death of Purkinje cells in the mouse cerebellum (Lee *et al.*, 2006a). In order to clarify if the *C. albicans* mistranslating

strains also accumulated insoluble proteins quantification of cytosolic protein aggregates was performed. Surprisingly, there was no significant difference between the control and mistranslating strains with the exception of the strain misincorporating serine at the leucine CTC codon (Figure 3-6). This result suggests that the *C. albicans* proteome is somehow pre-adapted to mistranslation. The unique exception of the strain misincorporating serine at the leucine CTC codon could be explained by its codon usage. The CTC codon is used at low level and possibly the ratio of serine misincorporation at this codon is higher, leading to production of higher number of erroneous proteins.

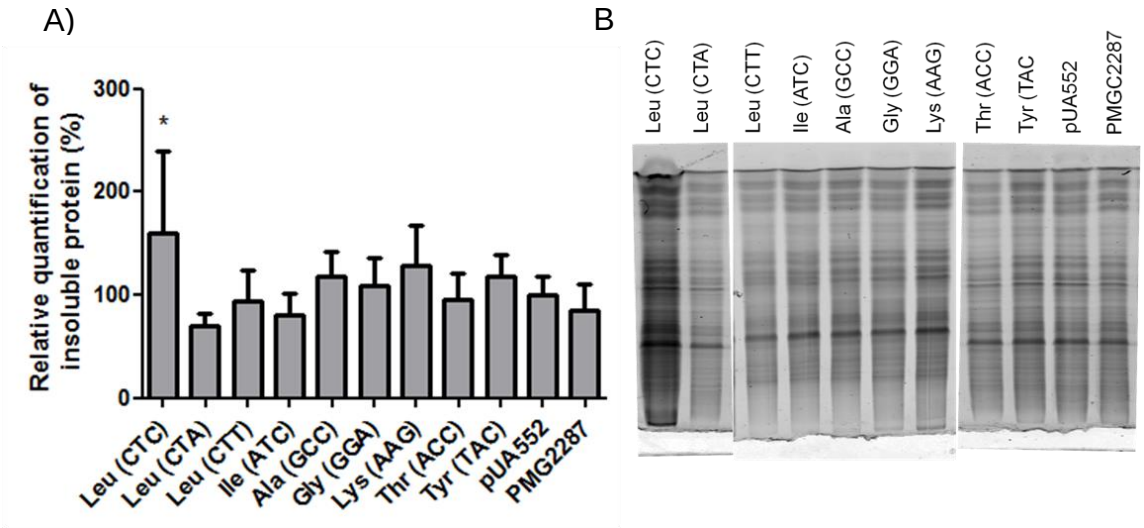


Figure 3-6: The effect of mistranslation on protein aggregation. **A)** The graph shows small increases in protein aggregation induced by mistranslation. The data represents insoluble protein + s.d. of duplicates of 3 different clones, normalized to the pUA552 control. Statistical analysis was carried out using one-way ANOVA was followed by a Dunnet test with CI 95% relative to pUA552 (*p<0.05). **B)** Gel showing insoluble protein fractionated on a 15% acrylamide gel.

Previous studies showed that despite proteome disruption, *S. cerevisiae* strains mistranslating the CUG codon have a competitive advantage under specific stress conditions, namely high temperature, salts, heavy metals and oxidants (Santos *et al.*, 1999). To clarify whether mistranslation of different codons of *C. albicans* would also produce positive outcomes in different stress conditions, a screen assay testing 21 different conditions was carried out. These tests probed a broad spectrum of metabolic networks that included growth with different carbon sources, at different pH values (pH 5–8.6), at different temperatures (25°C–42°C), on agar plates supplemented with elevated cation concentrations or antifungal

agents, or under other conditions that caused cellular stress (Table 3-3). For drug and nutrient concentrations were calibrated to allow for observation of both impairment and enhancement of growth relative to control. Independent isolates of each mistranslating strain were plated as a series of 10 fold dilutions on solid media using a 96-pin bolt replicator (Caliper). After 3-5 days of incubation, plates were photographed using Axio Vision Software from Zeiss. All images were imported and processed using ImageJ software and scored for growth phenotypes by comparison to the control strain (pUA552) included on the same plate. This approach generated individual growth scores for each mutant on each growth medium (Figure 3-7-A). Hence, the scoring system measured the reduction or enhancement of growth relative to control (pUA552).

The phenotypic profiles produced by the array above are provided in Figure 3-7. The colour scale represents the range of phenotype variability from strong enhancement of growth (red) to strong reduction of growth (green). The color intensity represents the strength of the phenotype and white represents a phenotype that is indistinguishable from the control strain pUA552. The majority of the strains showed growth disadvantages at 37ºC, except strains misincorporating serine at leucine CTA and alanine GCC codons. At 42ºC the strains misincorporating serine at lysine AAG and tyrosine TAC codons showed a slight growth advantage (Figure 3-7). Strains misincorporating serine at leucine CTC and CTA and threonine ACC codons displayed advantageous growth in media containing $CaCl_2$, $CuSO_4$ and galactose as single carbon source. Medium at pH 5.0 proved to be deleterious to strains misincorporating serine at leucine CTC and CTT, isoleucine ATC and alanine GCC codons. Although the changes in growth were idiosyncratic with each strain exhibiting a unique pattern of phenotypes, all strains showed an increased resistance to sorbitol (1.5M) and guanidine hydro chloride (5mM). When exposed to urea (25mM), strains misincorporating serine at leucine, threonine and tyrosine codons also displayed a growth advantage. Finally, the sensitivity to oxidative stress was tested in media containing hydrogen peroxide (H_2O_2), and the strains that misincorporated serine at leucine CTC, glycine GGA, lysine AAG and threonine ACC codons, also displayed a growth advantage.

Overall, there was variability between clones and stress conditions, nevertheless, misincorporating strains performed better than the control strain in numerous growth conditions indicating that despite the expected proteome disruption, the manipulation of the genetic code can have positive outcomes.

A)

Assay	Leu (CTC)	Leu (CTA)	Leu (CTT)	Ile (ATC)	Ala (GCC)	Gly (GGA)	Lys (AAG)	Thr (ACC)	Tyr (TAC)
25ºC	14.57	6.36	4.03	-4.02	8.11	-11.51	2.73	12.71	6.23
30ºC	37.96	19.4	1.74	50.46	-0.41	8.79	16.04	31.82	44.56
37ºC	-70.62	20.9	-17.7	-35.5	0.5	-23.8	-16.4	-61.61	-27.3
42ºC	-24.08	-30.22	-11.94	-39.22	-29.77	-19.59	3.5	-34.07	12.1
CaCl2 300 mM	257	100.4	6.2	-1.58	25.8	-5.69	20.6	160.2	43.5
NaCl 1.3 M	46.11	-1	5.22	6.92	17.14	1.51	3.24	18.54	2.16
C-absent	-12.54	-11.09	0.42	1.24	0.87	-9.95	-5.05	-2.65	-5.11
C-Galactose	173.7	89.9	13.1	39	17	2.9	10.6	58	14.9
C-Glycerol	-11.7	-4.27	6.46	-11.48	-8.73	1.12	11.15	-4.65	-1.42
C-Ethanol	-0.47	-4.96	3.48	1.82	1.98	-8.93	-5.49	-11.08	-0.51
Fluconazole 0.5 µg/ml	30.95	1.02	25.23	-1.52	7.98	21.84	18.96	-5.76	15
Calcofluor White 20 µM	114.9	19.6	3.5	-7.4	-5	-11	7.7	62.8	10.4
Sorbitol 1.5 M	5.24	94.29	10.98	15.04	24.26	13.06	15.33	36.49	24.52
Caffeine 15 mM	-0.01	-0.18	3.73	4.93	4.95	5.62	2.67	3.52	5.83
EDTA 0.75 mM	-23.47	-3.56	-10.25	31.07	-1.07	36.37	8.1	8.73	58.07
Hydrogen Peroxyde 6.0 mM	18.58	-4.66	-4.32	4.75	-3.9	32.98	26.48	27.18	69.68
pH 5.0	-94.33	8.57	-32.02	-46.45	-51.74	-7.26	21.27	4.77	-6.65
pH 8.6	48.2	-27.1	-6.9	17	-32.61	-18.9	-3.4	42.3	12.1
CuSO4 13 mM	190.02	59.24	-8.16	2.88	93.12	44.52	-6.5	174.72	15.78
Guanidine HCl 5 mM	19.66	38.96	35.06	18.16	23.86	10.66	2.8	18.36	8.36
Urea 25 mM	54.3	26.4	12.7	3.8	-10.8	-0.8	11.8	41.7	38.3

B)

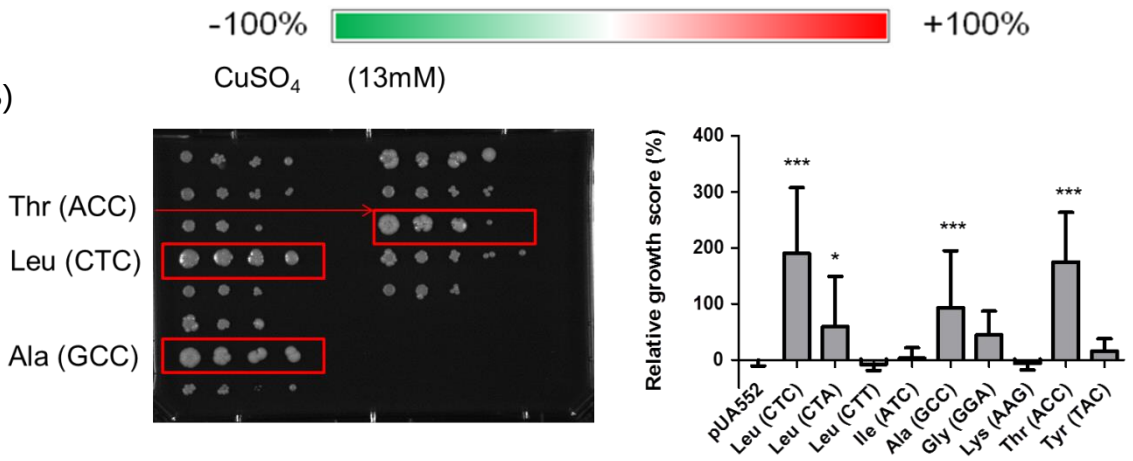


Figure 3-7: Phenotypic assay result table. A) Data represent the average size of the colonies obtained for all the transformed cells after 5 days of growth under the indicated conditions normalized with growth at 30ºC and compared with the normalized control pUA552 (considered 0%) for that condition (relative growth score). The assay was performed with two independent growth cultures and three independent clones. Brick color intensity represents the strength of the phenotype. White squares represent a phenotype that is indistinguishable from the control strain. Green and red squares represent either a growth reduction or enhancement, respectively. **B)** Growth on YEPD supplemented with 13 mM of CuSO4. Strains misincorporating serine in leucine CTC, alanine GCC and threonine ACC showed an advantageous growth phenotype. Data represent the mean ± standard deviation of triplicates of 3 independent clones (***p<0.001; **p<0.01; *p<0.05 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the pUA552 control cells).

C. albicans exposed to stress, namely antifungals, DNA transformation, growth in the host or heat shock, can undergo genomic alterations. In particular, chromosomal rearrangements, aneuploidy, and loss of heterozygosity (LOH) events are common (Bouchonville *et al.*, 2009;Selmecki *et al.*, 2010). These alterations are thought to control the expression of catabolic pathways, playing an important role in survival and adaptation of microorganisms (Rustchenko, 2007). Since mistranslation is usually toxic, we decided to investigate its impact on DNA content using flow cytometry. *C. albicans* exists as a diploid (2N) yeast, however some clones often become polyploid (Scherer and Magee, 1990). In our study, strains misincorporating serine at leucine CTC, CTA codons showed a shift in ploidy from 2.0N to 2.50 and 2.54N, respectively (Figure 3-8).

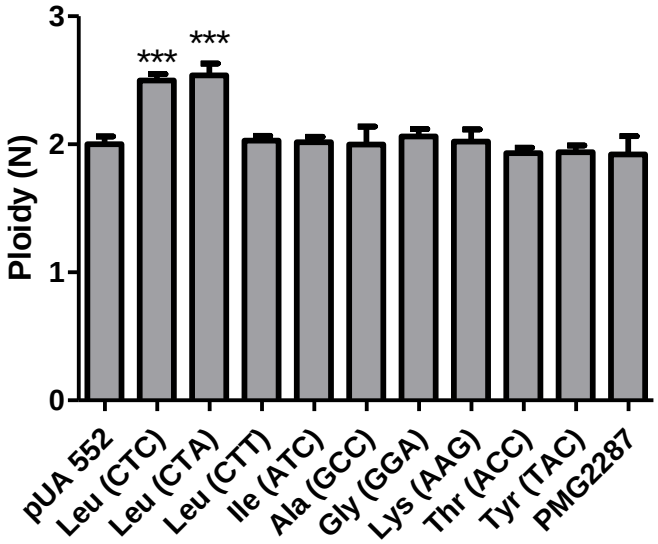


Figure 3-8: The effect of mistranslation on yeast ploidy. The plots represent the mean of the ploidy + s.d. of 3 different clones, normalized to the pUA552 control. Strains misincorporating serine at threonine codons (Thr) and Tyrosine codons (Tyr) are represented only for 2 different clones the remaining clones are described in Figure 3-9. Statistical analysis of the data was carried out using a one-way ANOVA was performed followed by a Dunnet test with CI 95% relative to pUA552 (***p<0.001).

One of the 3 clones of the strain that misincorporated serine at threonine codons showed a heterogeneous population of tetraploid and diploid cells (Figure 3-9-A). On the other hand, one clone misincorporating serine at tyrosine codons showed a heterogeneous population and cells with lower ploidy were 1.59 N

(Figure 3-9-B). These ploidy shifts were previously observed in *C. albicans* cells misincorporating serine at the leucine CUG codon, where heterogeneous populations of cells with 2, 4, 6 and 8 N were observed (Miranda *et al.*, 2007).

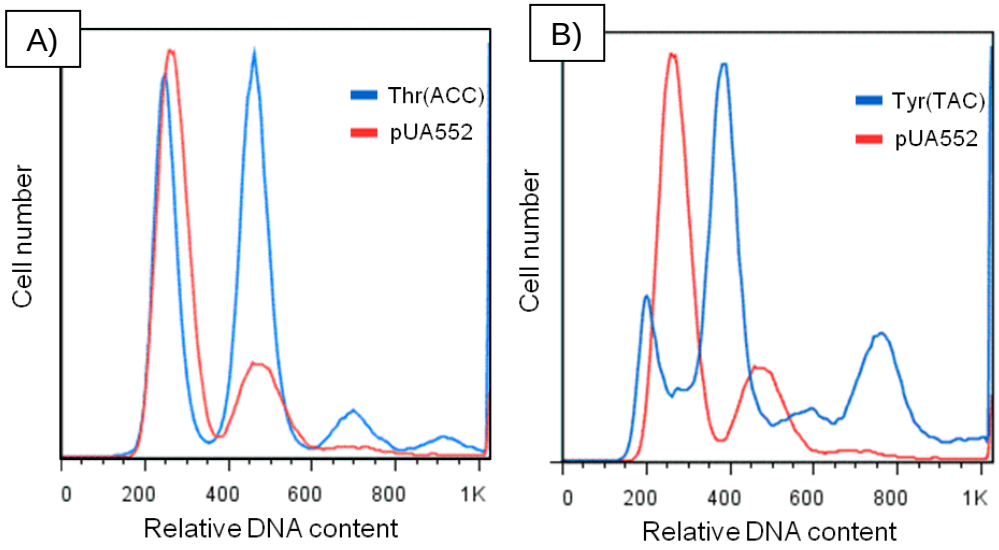


Figure 3-9: FACS analysis of the DNA content of the control (pUA552) and strains misincorporating serine at threonine and tyrosine codons. A) Strain misincorporating serine at threonine codons (blue line) showed 4 peaks corresponding to 2, 4, 6 and 8N**. B)** Strain misincorporating serine at tyrosine codons showed 4 peaks, but when (blue line) compared with control (red line) its ploidy of signal (200) is 1.59N.

3.5. Discussion

The constructed chimeric tRNAs and the misincorporation of serine at different amino acid sites proved to be a challenge for *C. albicans*. The low transformation efficiency (Figure 3-1) and growth rates (Figure 3-3) indicate that the engineered serine misincorporations are toxic. This toxicity is probably related with the wide-proteome mutagenesis caused by the substitution of an amino acid with serine. The different properties of the targeted amino acids likely resulted in production of gain-of-function (GOF) or misfolded proteins (Geslain *et al.*, 2010). Besides the different chemical characteristics of the amino acids, other important variables were the usage of the targeted codons and the localization of the misincorporated amino acid in the proteins. Generally, the amino acids tryptophan,

methionine, phenylalanine, isoleucine, valine, leucine, alanine and tyrosine are predominantly hydrophobic and are frequently buried in the proteins, while the amino acids threonine, serine, asparagine, glutamine, arginine, aspartate, glutamate and lysine are polar amino acids and are regularly exposed. Glycine and histidine occur equally buried or at protein surface (Fiser *et al.*, 1996). The chimeric tRNAs that reduced most transformation efficiency were those causing misincorporation of serine at leucine, isoleucine and threonine codons. The results of the misincorporation at leucine and isoleucine codons are in agreement with the observation from Geslain and colleagues, where isoleucine caused the strongest inhibition of cell division (Geslain *et al.*, 2010). However, the strain misincorporating serine at the threonine codon sites should not be the most affected, according to the amino acid substitution matrix BLOSUM 62 (Figure 3-10). This matrix compares the likelihood of substitution rate of an amino acid for another one based on multialignments of orthologous proteins. Replacement of threonine for serine has a substitution value of +1, which means that is a favored change (Henikoff and Henikoff, 1992) .

The transformation efficiency of the strain carrying the WT tRNA (pUA552) was also lower than that of the strain carrying the empty plasmid (PMG2287). However growth rate and ^{35}S-Met incorporation was higher than that of the empty plasmid (PMG2287). This increase of growth and protein synthesis rates can be linked with the overexpression of tRNA genes that occurs in high translation activity (Zhou *et al.*, 2009). Among the strains misincorporating serine at leucine codons, the smallest effect on growth rate occurred in strains whose codon were used at higher level, namely codons CTC, CTA and CTT (codon usage of 2.6, 4.4, and 10.2, respectively). It is therefore likely that the ratio between aminoacylated chimeric tRNA and the aminoacylated WT tRNA, which are competing for the same codon, are responsible for those effects. Since tRNA abundance is gene copy number dependent and the chimeric tRNA is expressed constitutively, a higher misincorporation of serine is expected in the codons with lower codon usage.

The phenotypic assay (Figure 3-7) revealed decreased growth for all strains at 37ºC except for strains misincorporating serine at leucine (CTA) and alanine

(GCC) codons. At 42ºC there was also a growth decrease except for strains misincorporating serine at lysine (AAG) and tyrosine (TAC) codons. This growth decrease is likely related with proteome disruption and heat shock proteins (Hsp) availability, since Hsp 70, Hsp 90 and Hsp 100 which are predominant in heat shock response (Burnie *et al.*, 2006). These Hsps are responsible for protecting proteins during thermal stress and for facilitating repair or degradation of damaged proteins (Panaretou and Zhai, 2008). The mistranslating strains likely produce large number of erroneous proteins which should sequester Hsps, preventing their rapid response to increased temperature. All strains showed increased resistance when exposed to sorbitol (1.5M) and guanidine hydrochloride (5mM). When exposed to urea (25mM), strains misincorporating at leucine, threonine and tyrosine codons also displayed a growth advantage. Guanidine hydrochloride and urea are protein denaturation agents and are expected a strong negative effect on growth of the misincorporating strains. However, this was not observed and strains grew faster than the control strain in presence of these chemicals. A possible explanation for these results is that mistranslation up-regulates genes that are involved in the response to sorbitol, urea and guanidine hydrochloride, protecting cells from these agents.

Alteration in DNA content was also observed in the mistranslating strains, namely in the strain misincorporating serine in leucine (CTC and CTA) codons (Figure 3-8), and in the strains misincorporating serine at tyrosine (TAC) and threonine (ACC) (Figure 3-9). Interestingly, these were the strains that performed better in the phenotypic assay. These results are in line with previous studies where chromosomal rearrangements, aneuploidy and loss of heterozygosity (LOH), played an important role in survival and adaptation of microorganisms by controlling the expression of catabolic pathways (Rustchenko, 2007;Bouchonville *et al.*, 2009;Selmecki *et al.*, 2010). A recent *C. albicans* study also showed that two extra copies of the left arm of chromosome 5, resulted in resistance to fluconazole, because of the higher copy number of the genes ERG11 (fluconazole target) and TAC1 a transcription factor (Selmecki *et al.*, 2006).

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Figure 3-10: Blosum 62 matrix for amino acids substitution. In the grey boxes are highlighted the amino acids that were targeted for substitution by serine. The values indicated are derived from amino acids pair frequencies at "blocks" of conserved sequences found in multiple protein alignments. The higher the indicated value, the more likely that amino acid is substituted. Adapted and modified from (Henikoff and Henikoff, 1992).

3.6. Conclusion

Our study shed new light in how artificial genetic code ambiguity affects the fitness of a eukaryote organism with an already ambiguous genetic code, as in the case of *C. albicans*. The artificial ambiguity was toxic to *C. albicans* strains, in particular to those misincorporating serine at leucine codons. Although, when these misincorporating strains were submitted to new stressful conditions, namely drugs, they revealed some adaption capacity. These ambiguous strains showed therefore that ambiguity could have played an important role in the evolution of the genetic code and also in adaptation to new environments.

3.7. References

Alamgir,M., Eroukova,V., Jessulat,M., Xu,J., and Golshani,A. (2008). Chemical-genetic profile analysis in yeast suggests that a previously uncharacterized open reading frame, YBR261C, affects protein synthesis. BMC. Genomics *9*, 583.

Ambrogelly,A., Palioura,S., and Soll,D. (2007). Natural expansion of the genetic code. Nat. Chem. Biol. *3*, 29-35.

Anderson,J.C., Wu,N., Santoro,S.W., Lakshman,V., King,D.S., and Schultz,P.G. (2004). An expanded genetic code with a functional quadruplet codon. Proceedings of the National Academy of Sciences *101*, 7566-7571.

Barelle,C.J., Manson,C.L., MacCallum,D.M., Odds,F.C., Gow,N.A., and Brown,A.J. (2004). GFP as a quantitative reporter of gene regulation in *Candida albicans*. Yeast *21*, 333-340.

Barrell,B.G., Bankier,A.T., and Drouin,J. (1979). A different genetic code in human mitochondria. Nature *282*, 189-194.

Bouchonville,K., Forche,A., Tang,K.E., Selmecki,A., and Berman,J. (2009). Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*. Eukaryot. Cell *8*, 1554-1566.

Burnie,J.P., Carter,T.L., Hodgetts,S.J., and Matthews,R.C. (2006). Fungal heat-shock proteins in human disease. FEMS Microbiol. Rev. *30*, 53-88.

Crick,F.H. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367-379.

Cropp,T.A. and Schultz,P.G. (2004). An expanding genetic code. Trends Genet. *20*, 625-630.

Di,G.M. (2005). The origin of the genetic code: theories and their relationships, a review. Biosystems *80*, 175-184.

Fiser,A., Simon,I., and Barton,G.J. (1996). Conservation of amino acids in multiple alignments: aspartic acid has unexpected conservation. FEBS Lett. *397*, 225-229.

Gasch,A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D., and Brown,P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241-4257.

Gasch,A.P. and Werner-Washburne,M. (2002). The genomics of yeast responses to environmental stress and starvation. Funct. Integr. Genomics *2*, 181-192.

Geslain,R., Cubells,L., Bori-Sanz,T., varez-Medina,R., Rossell,D., Marti,E., and de Pouplana,L.R. (2010). Chimeric tRNAs as tools to induce proteome damage and identify components of stress responses. Nucleic Acids Res. *38*, e30.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206.

Haig,D. and Hurst,L.D. (1991). A quantitative measure of error minimization in the genetic code. J. Mol. Evol. *33*, 412-417.

Henikoff,S. and Henikoff,J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A *89*, 10915-10919.

Homann,O.R., Dea,J., Noble,S.M., and Johnson,A.D. (2009). A phenotypic profile of the *Candida albicans* regulatory network. PLoS. Genet. *5*, e1000783.

Kano,A., Ohama,T., Abe,R., and Osawa,S. (1993). Unassigned or nonsense codons in Micrococcus luteus. J. Mol. Biol. *230*, 51-56.

Knight,R.D., Freeland,S.J., and Landweber,L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet *2*, 49-58.

Lee,J.W., Beebe,K., Nangle,L.A., Jang,J., Longo-Guess,C.M., Cook,S.A., Davisson,M.T., Sundberg,J.P., Schimmel,P., and Ackerman,S.L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature *443*, 50-55.

Lenhard,B., Orellana,O., Ibba,M., and Weygand-Durasevic,I. (1999). tRNA recognition and evolution of determinants in seryl-tRNA synthesis. Nucleic Acids Res. *27*, 721-729.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A Genetic Code Alteration Is a Phenotype Diversity Generator in the Human Pathogen *Candida albicans*. PLoS. ONE. *2*, e996.

Noble,S.M. and Johnson,A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. Eukaryot. Cell *4*, 298-309.

Oba,T., Andachi,Y., Muto,A., and Osawa,S. (1991). CGG: an unassigned or nonsense codon in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. U. S. A *88*, 921-925.

Ohama,T., Inagaki,Y., Bessho,Y., and Osawa,S. (2008). Evolving genetic code. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *84*, 58-74.

Panaretou,B. and Zhai,C. (2008). The heat shock proteins: Their roles as multi-component machines for protein folding. Fungal Biology Reviews *22*, 110-119.

Paredes,J.A., Carreto,L., Simoes,J., Bezerra,A.R., Gomes,A.C., Santamaria,R., Kapushesky,M., Moura,G.R., and Santos,M.A. (2012). Low level genome mistranslations

deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC. Biol. *10*, 55.

Pezo,V., Metzgar,D., Hendrickson,T.L., Waas,W.F., Hazebrouck,S., Doring,V., Marliere,P., Schimmel,P., and de Crecy-Lagard,V. (2004). Artificially ambiguous genetic code confers growth yield advantage. Proc. Natl. Acad. Sci. U. S. A *101*, 8593-8597.

Rand,J.D. and Grant,C.M. (2006). The thioredoxin system protects ribosomes against stress-induced aggregation. Mol. Biol. Cell *17*, 387-401.

Rustchenko,E. (2007). Chromosome instability in *Candida albicans*. FEMS Yeast Res. *7*, 2-11.

Santos,M.A., Cheesman,C., Costa,V., Moradas-Ferreira,P., and Tuite,M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida spp*. Mol. Microbiol. *31*, 937-947.

Santos,M.A., Gomes,A.C., Santos,M.C., Carreto,L.C., and Moura,G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607-611.

Santos,M.A., Perreau,V.M., and Tuite,M.F. (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. EMBO J. *15*, 5060-5068.

Santos,M.A. and Tuite,M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res. *23*, 1481-1486.

Santos,M.A., Ueda,T., Watanabe,K., and Tuite,M.F. (1997). The non-standard genetic code of *Candida spp*.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423-431.

Scherer,S. and Magee,P.T. (1990). Genetics of *Candida albicans*. Microbiol. Rev. *54*, 226-241.

Schmitt,M.E., Brown,T.A., and Trumpower,B.L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res. *18*, 3091-3092.

Selmecki,A., Forche,A., and Berman,J. (2006). Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science *313*, 367-370.

Selmecki,A., Forche,A., and Berman,J. (2010). Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot. Cell *9*, 991-1008.

Soll,D. and RajBhandary,U.L. (2006). The genetic code - thawing the 'frozen accident'. J. Biosci. *31*, 459-463.

Soma,A., Kumagai,R., Nishikawa,K., and Himeno,H. (1996). The anticodon loop is a major identity determinant of *Saccharomyces cerevisiae* tRNA(Leu). J. Mol. Biol. *263*, 707-714.

Srinivasan,G., James,C.M., and Krzycki,J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Walther,A. and Wendland,J. (2003). An improved transformation protocol for the human fungal pathogen *Candida albicans*. Curr. Genet *42*, 339-343.

Wang,L., Brock,A., Herberich,B., and Schultz,P.G. (2001). Expanding the genetic code of *Escherichia coli*. Science *292*, 498-500.

Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y., and Osawa,S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. U. S. A *82*, 2306-2309.

Zhou,Y., Goodenbour,J.M., Godley,L.A., Wickrema,A., and Pan,T. (2009). High levels of tRNA abundance and alteration of tRNA charging by bortezomib in multiple myeloma. Biochem. Biophys. Res. Commun. *385*, 160-164.

4. Transcriptional responses to genetic code ambiguity

4.1. Abstract

Naturally occurring genetic code alterations highlight the flexibility and evolvability of the genetic code (Knight *et al.*, 2001). *Candida albicans* and many other fungal species of the CTG clade translate CUG codons mainly as serine (Santos and Tuite, 1995;Suzuki *et al.*, 1997), using a novel ser-tRNA$_{CAG}$ (Santos *et al.*, 1993;Suzuki *et al.*, 1997). During the early stages of reassignment, the incorporation of serine in CUG codons was minimal, allowing the CTG clade ancestor to cope with ambiguous decoding. The levels of serine incorporation increased gradually while the incorporation of leucine decreased and eventually the gene encoding the leu-tRNA responsible for decoding CUG codons disappeared (Massey *et al.*, 2003;Miranda *et al.*, 2006). How the CTG clade ancestor tolerated the initial codon ambiguity is still unclear. To elucidate this question we have engineered *C. albicans* strains that ambiguously translate CTT (Leucine), ATC (Isoleucine), GCC (Alanine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) codons as serine. These strains were then characterized using DNA microarrays, which allowed us to obtain a global image of the cellular response to codon ambiguity. For comparative analysis, *Saccharomyces cerevisiae* strains mistranslating identical amino acids were also analyzed using DNA microarrays. Gene expression profiling of the *C. albicans* misincorporating strains grown at 30ºC deregulated 32.7% of its genes, while strains grown at 37ºC deregulated 18.8% of its genes. Transcriptome profiling of *S. cerevisiae* misincorporating strains showed deregulation of 43.7% of its genes. The only process that was up-regulated across all the strains tested was the oxidation reduction process. These transcriptome profiles also revealed that serine misincorporation at different codons generates different gene expression deregulations in different conditions. The overall data show major differences in the cellular response to codon ambiguity between *C. albicans* and *S. cerevisiae*, suggesting higher level of adaptation of the former to proteome disruption

4.2. Introduction

Variations to the standard genetic code have been reported in several organisms, totalizing 16 codon alterations in mitochondria and 10 codon alterations in eukaryotic nuclear and bacterial codes (Knight *et al.*, 2001). Several codons are more prone to change than others, i. e., the stop codons UAA, UAG and UGA, the AUA, AAA, AGA, AGG and CUG and the four-codon boxes CUN and CGN (Knight *et al.*, 2001;Di, 2005). In addition, the discovery of new natural proteinogenic amino acids such as selenocysteine (Ambrogelly *et al.*, 2007) and pyrrolysine (Srinivasan *et al.*, 2002;Soll and RajBhandary, 2006) further highlight variation, flexibility and evolvability of the genetic code. One of the organisms that altered the genetic code is the human pathogen *C. albicans*, which reassigned the CUG codon from leucine to serine (Santos *et al.*, 1997). Interestingly, this CUG codon is still ambiguously translated as serine (95-97%) and leucine (2.96% to 4.95%) (Gomes *et al.*, 2007), due to the hybrid nature of the $tRNA_{CAG}$, which is recognized by both the seryl- and leucyl-tRNA synthetases (SerRS and LeuRS) originating Ser-$tRNA_{CAG}^{Ser}$ and Leu-$tRNA_{CAG}^{Ser}$, respectively. Since Leu-$tRNA_{CAG}^{Ser}$ is not edited by the LeuRS nor discriminated by the translation elongation factor 1 (eEF1A) (Santos *et al.*, 1997;Santos *et al.*, 2011), these aminoacylated tRNAs compete for CUG codons during translation (Gomes *et al.*, 2007). The ambiguous translation of this CUG codon has been preserved over the last 272 ± 25 million years (My), for reasons that are not fully understood (Massey *et al.*, 2003). Current models of CUG reassignment postulate that the incorporation of Ser at CUG codons was low initially and increased gradually while the incorporation of Leu at CUG codons decreased (Mateus *et al.*, 2013). This lead to the disappearance of the gene that encoded the leu-tRNA responsible for the decoding CUG codons (Santos *et al.*, 1997;Massey *et al.*, 2003;Miranda *et al.*, 2006). This period of codon ambiguity must have been stressful to the CTG clade ancestor and should have reduced fitness (Massey *et al.*, 2003). Indeed, incorporation of Ser at Leu-CUG sites should have affected protein function and stability leading to the production of toxic aggregated proteins (Drummond and Wilke, 2009).

This raises the major question of how was the ambiguous $tRNA_{CAG}^{Ser}$ selected under negative pressure? Several recent studies show that increased mistranslation levels can be advantageous in specific ecological niches. For example, engineered *S. cerevisiae* cells mistranslating Leu-CUG as Ser upregulate the general stress response and heat shock proteins resulting in increased survival at 50ºC and tolerance to cadmium, arsenate and hydrogen peroxide (Santos *et al.*, 1999;Mateus *et al.*, 2013). Up-regulation of the stress response produces stress cross protection and allows yeast to survive in highly toxic environments (Santos *et al.*, 1997). In other words, adaptation to new ecological conditions is sustained by reprogramming gene expression and can be monitored using DNA microarrays (Gasch *et al.*, 2000;Gasch and Werner-Washburne, 2002). Several studies have also shown that transcriptome reprograming in *C. albicans* and *S. cerevisiae* exposed to heat shock, oxidative or reductive stress, osmotic stress, nutrient starvation, DNA damage and extreme pH oxidative stress is common and important for adaptation (Gasch *et al.*, 2000;Causton *et al.*, 2001;Enjalbert *et al.*, 2003;Enjalbert *et al.*, 2006;Wang *et al.*, 2006). These studies also show differential responses of *S. cerevisiae* and *C. albicans* to environmental stress. The *S. cerevisiae* environmental stress response (ESR) consists of 300 up-regulated and around 600 down-regulated genes (Gasch *et al.*, 2000;Causton *et al.*, 2001). This ESR is controlled by the transcription factors Msn2 and Msn4, which do not have a homologous functions in *C. albicans* (Nicholls *et al.*, 2004). The majority of the up-regulated genes are involved in carbohydrate metabolism, defense against reactive oxygen species, protein metabolism, intracellular signaling and DNA damage and repair (Gasch, 2007). On the other hand, *C. albicans* has a weaker ESR (Enjalbert *et al.*, 2003;Enjalbert *et al.*, 2006) controlled by the Hog1 (Enjalbert *et al.*, 2006).

The studies carried out by Gomes and Miranda exposed high tolerance of *C. albicans* to CUG ambiguity and high adaptation capacity of ambiguous cells (Gomes *et al.*, 2007;Miranda *et al.*, 2007), however it is not yet clear whether such tolerance is specific of the CUG codons or also exists in other codons. In other words, it is important to clarify whether *C. albicans* is tolerant to general codon ambiguity. To clarify this question we have engineered *C. albicans* strains to

misincorporate Ser at various non-cognate codons. For this, we have mutated a Ser tRNA anticodon and produced a series of misreading tRNAs that misincorporate serine at CTT (Leucine), GCC (Alanine), AAG (Lysine), ACC (Threonine) and TAC (Tyrosine) codons. The gene deregulation induced by these mistranslations was captured using DNA microarrays. The transcriptome profiling of the ambiguous *C. albicans* strains grown at 30ºC revealed deregulation of 32.7% of its genes, while strains grown at 37ºC deregulated only 18.8% of its genes. A comparative study using ambiguous *S. cerevisiae* grown at 30ºC showed deregulation of 43.7% of its genes. Although the majority of the deregulated genes were specific of each strain, genes involved in transport and oxidation reduction processes were horizontally deregulated. The overall data confirms the hypothesis that *C. albicans* is more tolerant to codon ambiguity than *S. cerevisiae*.

4.3. Material and Methods

4.3.1. Strain maintenance

4.3.1.1. *C. albicans*

The *C. albicans* strain SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434}) (Noble and Johnson, 2005) was a gift from Professor Alexander D. Johnson (University of California-San Francisco). SN148 was grown in YPD (1% yeast extract, 2% peptone, 2% dextrose) at 30ºC. Transformed strains were grown in synthetic minimal medium (0.67% yeast nitrogen base, 2% glucose, 0.2% Drop-out mix with all the essential amino acids). Engineered strains used in this study are described below.

Plasmid purification and construction

Plasmid pUA552 was constructed by PCR amplification of tDNAser (UGA) from *S. cerevisiae*, with primer oUA1671 5'-AAAGGTACCGAAGGAGGTGCAAGGGAAAAG-3' and primer oUA1672 5'-

TTTGGGCCCTCCGTGCATAACGAATGACTC-3' and was inserted between *KpnI* and *ApaI* restriction enzyme sites of PMG2287 (gift from Prof. Judith Berman). The tRNA sequence was mutated using the QuikChange Site-Directed Mutagenesis kit (Stratagene) with primers oUA1766 5'-TAAGGCGACAGACGTGAAATCTGTTGGGCTC-3' and oUA1767 5'-GAGCCCAACAGATTTCACGTCTGTCGCCTTA-3' with the purpose of exchanging T33 for G33 to decrease serylation efficiency of the tRNA (Santos *et al.*, 1996). Then plasmid pUA552 (tRNA anticodon) was mutated using the oligonucleotides described in Table 4-1 to produce new plasmids. These plasmids were linearized with *StuI* and integrated into the *RPS10* locus of the *C. albicans* genome.

Table 4-1: List of plasmids and primer sequences used to mutate the $tRNA^{Ser}$ anticodon. The new anticodons obtained by mutagenesis and corresponding codons and amino acids are also indicated

Plasmid	Primer Name	Primer Sequence (5'->3')	Anticodon (5'->3')	Codon (5'->3')	Amino acid
pUA535	oUA1723	TAAGGCGACAGACGAAGAATCTGTTGGGCTC	AAG	CTT	Leucine
	oUA1724	GAGCCCAACAGATTCTTCGTCTGTCGCCTTA			
pUA537	oUA1727	TAAGGCGACAGACGGGCAATCTGTTGGGCTC	GGC	GCC	Alanine
	oUA1728	GAGCCCAACAGATTGCCCGTCTGTCGCCTTA			
pUA542	oUA1747	TAAGGCGACAGACGCTTAATCTGTTGGGCTC	CUU	AAG	Lysine
	oUA1748	GAGCCCAACAGATTAAGCGTCTGTCGCCTTA			
pUA544	oUA1751	TAAGGCGACAGACGGGTAATCTGTTGGGCTC	GGU	ACC	Threonine
	oUA1752	GAGCCCAACAGATTACCCGTCTGTCGCCTTA			
pUA546	oUA1755	TAAGGCGACAGACGGTAAATCTGTTGGGCTC	GUA	TAC	Tyrosine
	oUA1756	GAGCCCAACAGATTTACCGTCTGTCGCCTTA			

C. albicans transformation

Transformation of *C. albicans* was carried out using an improved lithium acetate method (Walther and Wendland, 2003) with small modifications. Briefly, overnight cultures were inoculated in fresh medium to an $O.D_{600nm}$ of 0.3 and let grow at 30ºC with 200rpm shaking until an $O.D_{600nm}$ of 1-1.2. Cell cultures were then transferred to 50ml falcon tubes and centrifuged at 4000rpm for 5min,

supernatants were discarded and pellets were resuspended in 1.5ml of LiAc-solution (0.1M LiAc, TE buffer 1x). 200 µl of suspension were transferred to 1.5mL eppendorf tubes and the transformation reagents were added in the following order: 600µL 50% (w/w) PEG LiAc-solution, 50µL single-stranded carrier DNA (2mg/mL) previously denatured and 50µL of an aqueous solution of the plasmid of interest (containing 1– 5µg of plasmid). Tubes were vortexed and incubated at 30ºC for 4 hours following incubation at 44ºC for 15min and then placed on ice for 2min. Cells were harvested at 4000rpm for 5min and resuspended gently in 200µl of appropriate SC selective medium. Each 100µl of suspension was plated onto appropriate SC selection medium plates and incubated at 30°C for 3-4 days.

4.3.1.2. S. cerevisiae

The *S. cerevisiae* BY4743 strain (*MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0)* was used to engineer the yeast misincorporating strains (Mateus, 2011). The plasmids used to transform BY4743 are indicated in Table 4-2. Transformed strains were grown in synthetic minimal medium; 0.67% yeast nitrogen base, 2% glucose, 0.2% drop-out mixture with all the essential amino acids.

Table 4-2: Plasmid description of *S. cerevisiae* mistranslating strains. The corresponding codons and amino acids are also indicated.

Plasmid	Descriptrion	Codon (5'->3')	Amino acid
pUA261	pRS315 plasmid containing one copy of *C. albicans* $tRNA_{UGA}^{Ser}$ gene inserted between *BamHI* and *SalI* restriction sites.	TCA	Serine
pUA262	Plasmid based on pUA261. Were *C. albicans* $tRNA_{UGA}^{Ser}$ gene was mutated to *C. albicans* $tRNA_{CAG}^{Ser}$	CTG	Leucine
pUA263	Plasmid based on pUA261. Were *C. albicans* $tRNA_{UGA}^{Ser}$ gene was mutated to *C. albicans* $tRNA_{GUA}^{Ser}$	TAC	Tyrosine
pUA264	Plasmid based on pUA261. Were *C. albicans* $tRNA_{UGA}^{Ser}$ gene was mutated to *C. albicans* $tRNA_{UUU}^{Ser}$	AAA	Lysine
pUA265	Plasmid based on pUA261. Were *C. albicans* $tRNA_{UGA}^{Ser}$ gene was mutated to *C. albicans* $tRNA_{CGU}^{Ser}$	ACG	Threonine
pUA268	Plasmid based on pUA261. Were *C. albicans* $tRNA_{UGA}^{Ser}$ gene was mutated to *C. albicans* $tRNA_{UGC}^{Ser}$	GCA	Alanine

4.3.3. Total RNA extraction

About 25 $O.D_{600nm}$ units (volume of culture X OD_{600nm}) of exponentially growing cells (OD_{600nm} = 0.5~0.8) were harvested by centrifugation at 4000 rpm for 4 minutes at room temperature. After removal of the supernatant, tubes were immediately immersed in liquid nitrogen for 6 minutes, frozen and stored at -80ºC. Total RNA was isolated from yeast using a hot phenol based protocol (Schmitt *et al.*, 1990b). Frozen pellets were resuspended in 500μl acid phenol chloroform (Sigma, 5:1, pH 4.7) at 65ºC. An equal volume of TES-buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% sodium dodecyl sulphate) was added and each tube was vortexed at high speed for 20 seconds to resuspend the pellets. Tubes were then incubated for 1 hour in a water bath at 65ºC with 20 seconds vortexing every 10 minutes. Tubes content were transferred to 1.5 ml tubes and centrifuged for 20 minutes, at 14000 rpm, at 4ºC. The water-phase was added to a new tube, filled with 600 μl Acid Phenol Chloroform (Sigma, 5:1, pH 4.7), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm at 4ºC. For a second extraction the water-phase from this step was added to new Eppendorf tubes, filled with 500 μl Chloroform:Isoamyl-alcohol (Sigma, 25:1), vortexed for 20 seconds and centrifuged, for 10 minutes, at 14000 rpm, at 4ºC. The water-phase was transferred again to a new Eppendorf tube filled with 400 μl Chloroform:Isoamyl-alcohol (Sigma, 25:1), vortexed for 20 seconds, and centrifuged for 10 minutes, at 14000 rpm, at 4ºC. Finally, the water-phase was added to new Eppendorf tubes containing 35 μl of sodium acetate (3 M, pH 5.2), 800 μl ethanol (100 %, kept at – 20ºC) and incubated at – 20ºC for at least 1 hour. Tubes were then centrifuged for 5 minutes at room temperature, 14000 rpm, and supernatants were removed carefully with a pipette avoiding touching the RNA-pellet. Pellets were then washed with 500 μl of ethanol (80 %, -20ºC) and centrifuged for 3 minutes at room temperature, 14000 rpm. Ethanol was removed, the RNA pellets were air dried for 1 minute and dissolved in sterile (mQ) water to a concentration of 10 μg / μl. Samples were stored at -80ºC.

4.3.4. DNA microarray

4.3.4.1 Labeling

DNA microarray experiments were carried out following protocols supplied by Agilent for One-Color Microarray-Based Gene Expression Analysis Low Input Quick Amp Labeling version 6.5 (Agilent Technologies). Labeling reactions were prepared using 200 ng of total RNA which was added to a spike control mixture. The cDNA was synthesized using T7 polymerase and poly (dT) primers. Labelling was done with the cyanine dye (Cy3-CTP), the amplified cRNA was then isolated using RNAeasy column purification kits (QIAGEN). Dye incorporation and cRNA yield were quantified using a NanoDrop ND-1000 Spectrophotometer. Slides were prepared using Agilent gasket slides according to the manufacturer instructions. A total of 1.65 µg of *C .albicans* samples and 0.825 µg of *S. cerevisiae* samples of Cy3-labeled cRNA were hybridized to appropriate DNA microarrays, using Agilent gasket slides according to the manufacturer instructions. Hybridizations were incubated for 17 h at 65° in a rotating oven. Following hybridization, microarrays were washed using the wash procedure with stabilization and drying solution (Agilent). Slides were then scanned using an Agilent G2565AA microarrays scanner (Agilent). Signal from probes features were extracted from the microarray scans data sets using Feature Extraction Software (Agilent).

4.3.4.2 Normalization

Extracted data were normalized using limmaGUI software (Bioconductor R) (Gentleman *et al.*, 2004). Normalized data from mistranslating strains were compared against normalized data from control strains using a t-test, *p*-value threshold of 0.05. The Gene Ontology enrichment analysis was done using genes with a fold change >1.5 or <-1.5. The *Saccharomyces* Genome Database (http://www.yeastgenome.org/) and *Candida* Genome Database (http://www.candidagenome.org/) were used for functional interpretation of the data and Enrichment of Gene Ontologies.

4.3.4.3. Transcription factor motifs

The detection of transcription motifs was performed using fasta files alignments 1000bp of sequence upstream of the AUG start codon of open reading frames (ORFs) that was submitted to the YEASTRACT online tool motive finder option “Find TF Binding Site(s)” (http://www.yeastract.com) (Monteiro *et al.*, 2008). The selected genes for these analyses were deregulated in at least 3 strains of *C. albicans* and 4 strains of *S. cerevisiae*. The transcription factor ratio obtained for each condition represents the number of genes where that transcription factor motif was detected relative to the total gene number tested for the respective condition and carried out in this analysis. As control the detection of transcription factor motifs was also carried out in 100 random genes. All the transcription factor motifs occurrences obtained were relative to *S. cerevisiae* transcription factor motifs. The correspondence to *C. albicans* transcription factor motifs was performed using the list of *S. cerevisiae* and *C. albicans* transcription factor orthologs constructed by Homann and coworkers (Homann *et al.*, 2009).

4.4. Results

To construct ambiguous strains a $tRNA_{UGA}^{Ser}$ gene was selected for mutational analysis and expression in yeast. For this, the anticodon nucleotides were mutated to produce tRNAs that could misincorporate Ser at various codons sites. Since the *C. albicans* SerRS identifies $tRNA^{ser}$ without recognizing the anticodon (it uses the acceptor stem, D-arm and extra stem/loop nucleotides to aid in the correct tRNA selection (Lenhard *et al.*, 1999)), mutant tRNAs retain their serylation capacity, leading to Ser insertion during ribosome translation of the codons complementary to the engineered anticodons. The Ser-chimeric tRNAs compete with the native aminoacylated tRNAs resulting in stochastic misincorporation of Ser at targeted codons.

4.4.1. Transcriptional responses to general gene mistranslations

To compare the *C. albicans* and *S. cerevisiae* responses to codon ambiguity, transcriptome profiling was performed in strains misincorporating Ser at leucine, alanine, lysine, threonine and tyrosine codons.

The strain with higher level of deregulated genes misincorporated Ser at Leu codons (Figure 4-1), but overall, gene deregulation was affected depending on strain (type of misincorporation) and growth condition. *C. albicans* showed higher number of deregulated genes at 30ºC than at 37ºC, in particular the strains misincorporating Ser at Leu and Ala codons (Figure 4-1 A-C).

The number of overlapping deregulated genes observed in different strains grown in the same media was lower in *C. albicans* (Figure 4-1, B and D). Surprisingly, deregulated genes were not shared between the *C. albicans* misincorporating strains, as each strain displayed a specific gene deregulation profile. Conversely, *S. cerevisiae* strains showed a common response (Figure 4-1 E) involving 99 commonly deregulated genes, and higher number of deregulated genes relative to *C. albicans.*

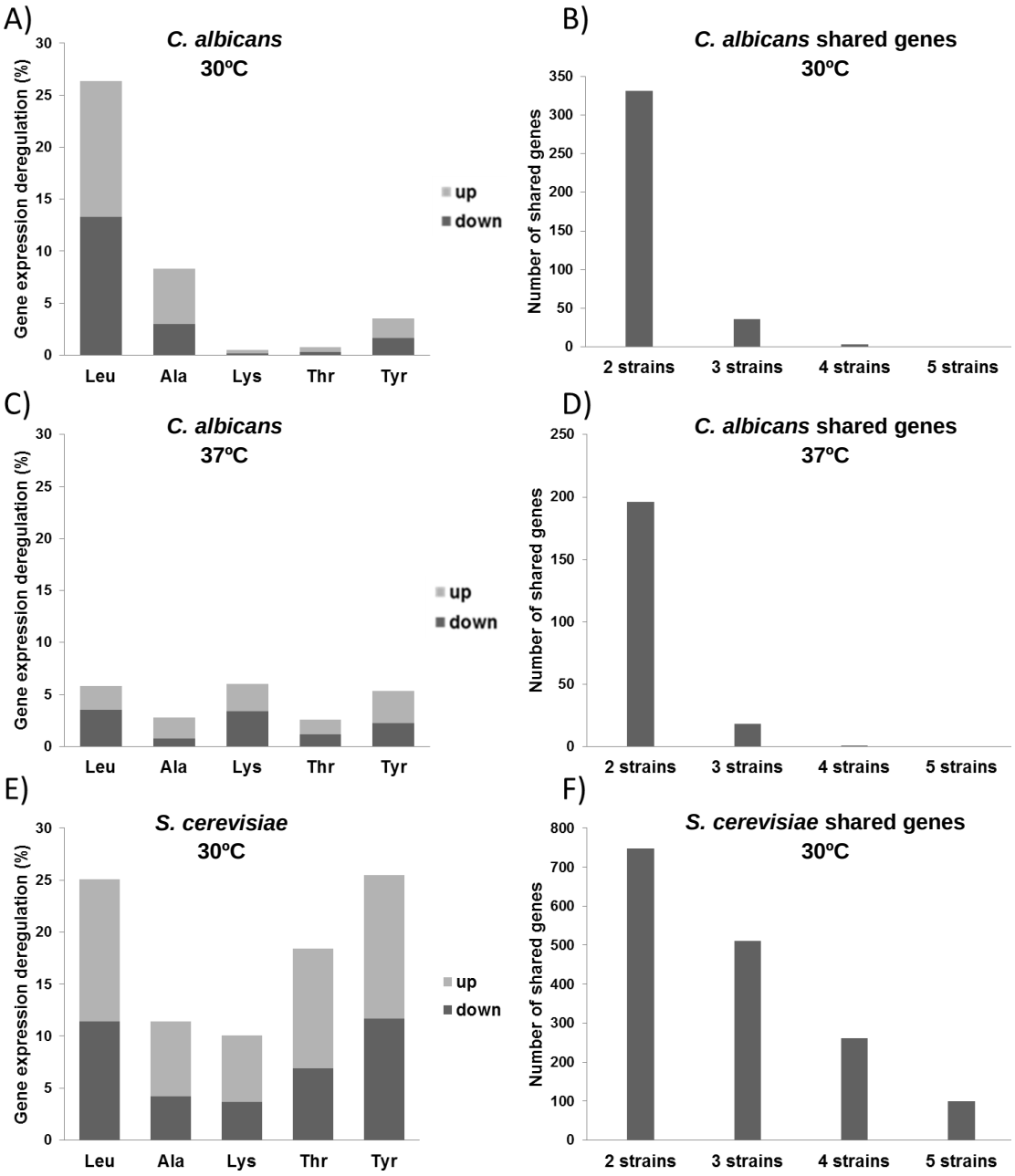


Figure 4-1: Gene expression deregulation in ambiguous *S. cerevisiae* and *C. albicans*. A) Percentage of deregulated genes in *C. albicans* grown at 30ºC, **B)** Total number of deregulated genes shared by *C. albicans* strains grown at 30ºC. **C)** Percentage of deregulated genes in *C. albicans* grown at 37ºC, **D)** Total number of deregulated genes shared by ambiguous strains of *C. albicans* grown at 37ºC. **E)** Percentage of deregulated genes of ambiguous *S. cerevisiae* grown at 30ºC, **F)** Total number of genes deregulated shared by ambiguous *S. cerevisiae* strains grown at 30ºC. The x-axis corresponds to misincorporation of Ser at codons belonging to the shown amino acids.

To understand if the observed transcriptional reprogramming was dependent on the amino acid that was replaced by Ser, we have compared the list

of *C. albicans* and *S. cerevisiae* deregulated genes for each amino acid in each growth condition. For this, a list of orthologous genes established by Enjalbert and co-workers (Enjalbert et al., 2006) was used and the overlapping genes were identified. There were no overlapped genes in the majority of the misincorporating strains (annexes table 1-10). The exception was the *C. albicans* strain misincorporating Ser at Leu codons that showed a small number of overlapping genes when grown at 30ºC and 37ºC with the *S. cerevisiae* strain also misincorporating Ser at Leu codons (Figure 4-2).

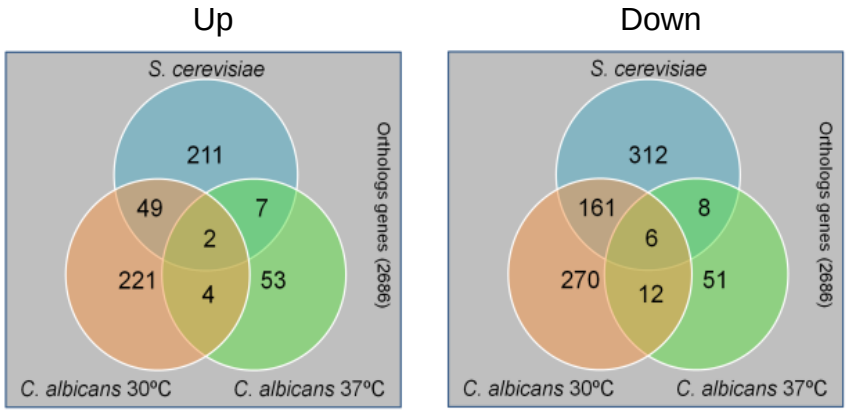


Figure 4-2: Venn diagrams showing gene deregulated by strains misincorporating Ser at Leu codons .The orthologs genes list used was previously described by Enjalbert (Enjalbert *et al.*, 2006). All genes shown have p-value <0.05. (Complete table in annexes, table 3 and 4).

The up regulated genes that overlapped in the three situations tested were *COX13* and *FBP1*. The former encodes the subunit VIa of cytochrome c oxidase and may minimize ROS production (Vukotic *et al.*, 2012). *FBP1* encodes fructose-1,6-bisphosphatase and its expression is important for the control of gluconeogenesis, but it is also involved in the response to aging and DNA damage (Kitanovic and Wolfl, 2006). Overlapping repressed genes were *NIP1*, *RPB9*, *YRB1*, *TOP2*, and *NOP15* that are involved in DNA and RNA metabolic process. *NOP15* was the most down-regulated gene and is involved in 60S ribosomal subunit biogenesis (Harnpicharnchai *et al.*, 2001). These results are in agreement with previous studies where *S. cerevisiae* misincorporating Ser at Leu CUG codons showed repression in ribosome biogenesis (Paredes *et al.*, 2012). The other shared downregulated gene was the *C. albicans MAK10.3* and the

corresponding *S. cerevisiae MAK10* that are involved in N-terminal protein amino acid acetylation processes (Polevoda *et al.*, 2009)

To compare the effects of mistranslation on the transcriptome of the strains used in this study, in particular to determine if *S. cerevisiae* and *C. albicans* responded similarly to mistranslation, hierarchically clustering of orthologs genes was carried out (Enjalbert et al., 2006).

The process that was most negatively affected by mistranslation was ribosome biogenesis and RNA processing. However, this effect was only observed in the *C. albicans* strain that misincorporated Ser at the Leu CTT codons, grown at 30ºC, and in the *S. cerevisiae* strains that misincorporated Ser at Leu, Thr and Tyr codons (Figure 4-3). Amino acid biosynthetic processes were up regulated in several strains grown at 30ºC, namely *C. albicans* that misincorporated Ser at Ala codons and *S. cerevisiae* that misincorporated Ser at Leu, Thr and Tyr codons, but those processes were down regulated in the *C. albicans* strains that misincorporated Ser at Lys and Thr codons (at 37ºC). Finally, the genes belonging to oxidation-reduction processes were up-regulated in *S. cerevisiae* strains that misincorporated Ser at Leu, Thr and Tyr codons. The hierarchical clustering of the mistranslating strains showed that the effect of mistranslation on gene deregulation segregated growth temperature, and *C. albicans* and *S. cerevisiae*, indicating that these variables are more important than the amino acid misincorporated.

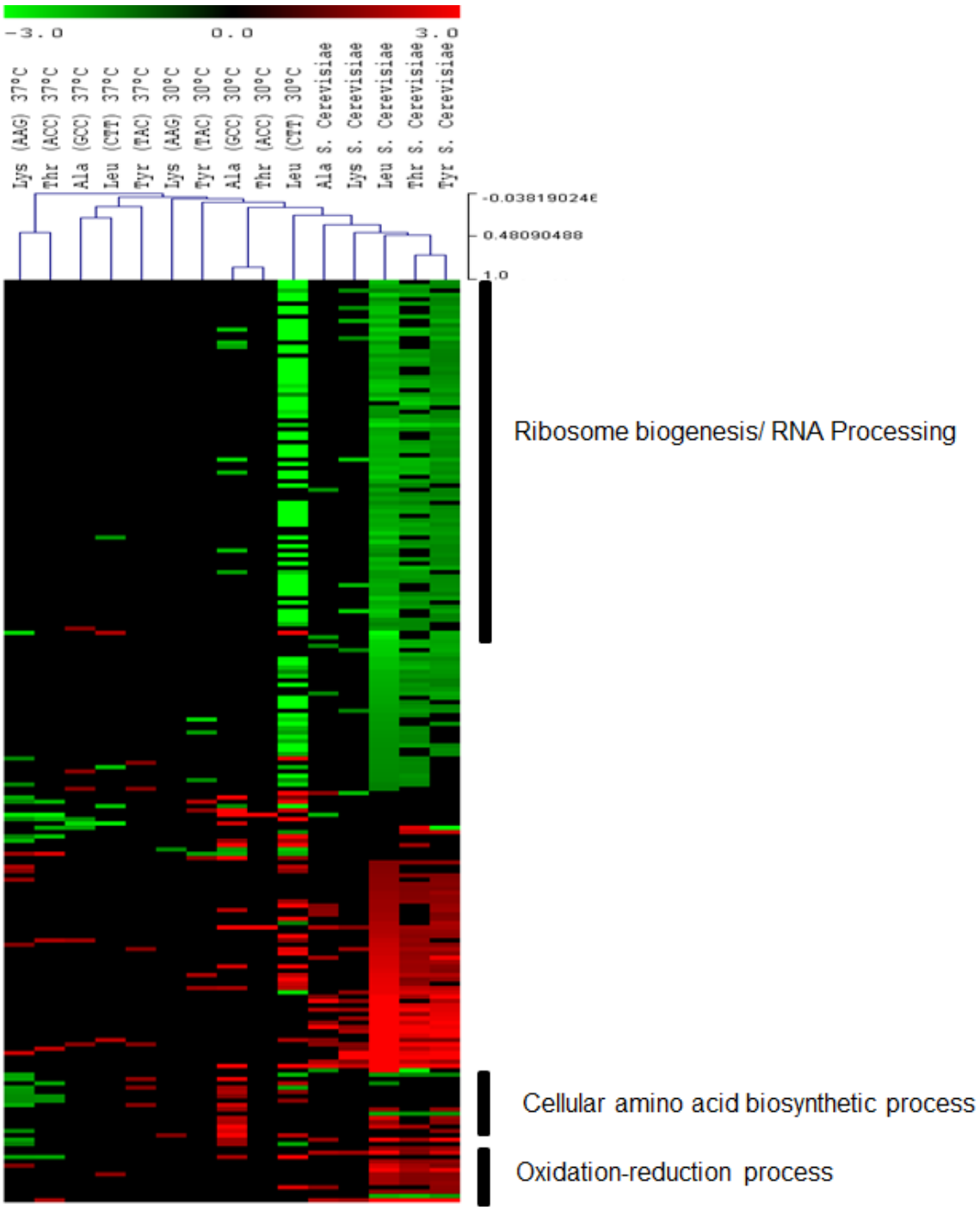


Figure 4-3: Clustering of the genes whose expression was deregulated by misincorporation of serine at alanine, leucine, lysine, threonine and tyrosine codons. Gene expression values shown are the normalized M-values against the corresponding control strain, using Welch approximation and p-value<0.05. Only genes with M-values corresponding to fold change <-1.5 and >+1.5 in at least 3 of the strains studied are shown. Strains are clustered hierarchically.

4.4.2. Functional class enrichment analysis

The enrichment analyses of down-regulated genes from *C. albicans* strains grown at 30ºC (Table 4-3) showed that gene expression, ribosome biogenesis, ribosome processing and translation, were the main cellular process affected by mistranslation. Interestingly, there was significant variability between strains and most of them were not affected by mistranslation at least in terms of the functional classes mentioned above. Similarly, enrichment analysis of up-regulated genes showed deregulation of oxidation reduction processes in the strains misincorporating Ser at Ala, Leu and Lys codons (Table 4-4). Interestingly, the former strain showed more extensive gene deregulation than the other strains despite the fact that Ser and Ala are chemically similar amino acids.

Table 4-3: Functional class enrichment analysis of genes down-regulated by mistranslation in *C. albicans* strains grown at 30ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+1.5 and <-1.5. (Complete table in annexes, table 11)

GOID	GO term	Genes	Ala		Leu		Lys		Thr		Tyr	
		in term	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6396	RNA processing	375			95	1.09E-17						
6412	translation	412									14	4.29E-03
10467	gene expression	971			153	5.15E-09					24	6.70E-04
42254	ribosome biogenesis	259	11	2.28E-02	90	1.29E-27						
44249	cellular biosynthetic process	1268									24	6.10E-02
90304	nucleic acid metabolic process	882			159	4.83E-15						

Table 4-4 : Functional class enrichment analysis of genes up-regulated by mistranslation in *C. albicans* strains grown at 30ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+1.5 and <-1.5. (Complete table in annexes, table 12).

GOID	GO term	Genes	Ala		Leu		Lys		Thr		Tyr	
		in term	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6520	cellular amino acid metabolic process	225	41	2.36E-16								
55085	transmembrane transport	329	39	4.40E-09								
55114	oxidation-reduction process	450	33	1.66E-02	74	4.30E-04	3	9.94E-03				

Regarding the enrichment analysis performed with the genes that were down regulated in the *C. albicans* strains grown at 37ºC (Table 4-5), the only process that overlapped was transmembrane transport. In this case, only strains mistranslating Ala and Thr codons showed down-regulation of this process. The

enrichment analysis of up-regulated genes (Table 4-6) showed that glucose catabolic processes were deregulated in strains mistranslating Ala and Tyr codons, respiratory electron transport chain processes were deregulated in strains misincorporating Ser at Leu and Tyr codons, and finally generation of precursor and metabolites and energy processes were deregulated in the strains misincorporating Ser in Ala, Leu and Tyr codons.

Table 4-5 : Functional class enrichment analysis of genes down-regulated by mistranslation in *C. albicans* strains grown at 37ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+ 1.5 and <-1.5. (Complete table in annexes, table 13)

GOID	GO term	Genes	Ala		Leu		Lys		Thr		Tyr	
		in term	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6520	cellular amino acid metabolic process	225					4	5.87E-03				
35672	oligopeptide transmembrane transport	6							2	8.16E-02		
44271	cellular nitrogen compound biosynthetic process	270					24	5.44E-03				
44281	small molecule metabolic process	665							2	8.16E-02		
55085	transmembrane transport	329	5	4.03E-02					8	9.55E-02		
55114	oxidation-reduction process	450							8	9.55E-02		

Table 4-6 : Functional class enrichment analysis of genes up-regulated by mistranslation in *C. albicans* strains grown at 37ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+ 1.5 and <-1.5. (Complete table in annexes, table 14)

GOID	GO term	Genes	Ala		Leu		Lys		Thr		Tyr	
		in term	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6091	generation of precursor metabolites and energy	130	8	6.29E-02	13	8.35E-09					11	1.31E-05
6412	translation	412	31	1.28E-14								
22904	respiratory electron transport chain	33			8	4.00E-08					7	6.93E-06
42254	ribosome biogenesis	259	15	1.70E-04								
55114	oxidation-reduction process	450			20	7.39E-08						

The functional enrichment of genes down-regulated in the *S. cerevisiae* strains misincorporating Ser at Leu, Thr and Tyr codons highlighted major changes of ribosomal biogenesis and assemble processes (Table 4-7). Previous studies with a mistranslation inducible *S. cerevisiae* strain revealed an identical trend concerning the decrease in the ribosomal biogenesis and assembly processes (Paredes *et al.*, 2012). Regarding the up-regulated genes (Table 4-8), a high number of them belong to the glycogen and trehalose metabolic processes, in particular in strains misincorporating Ser in Leu, Thr and Tyr codons. Oxidation

reduction processes enrichment was observed in the majority of strains except in strains mistranslating Ala codons.

Table 4-7 : Functional class enrichment analysis of genes down-regulated by mistranslation in *S. cerevisiae* grown at 30ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+1.5 and <-1.5. (Complete table in annexes, table 15)

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6396	RNA processing	567			155	2.72E-73			70	8.05E-33	78	9.94E-23
8033	tRNA processing	102			26	6.51E-09			11	6.53E-03		
10467	gene expression	1847			223	7.89E-41			98	2.82E-18		
42254	ribosome biogenesis	405			177	2.65E-127	13	8.74E-03	78	1.06E-52	96	6.30E-51
71826	ribonucleoprotein complex subunit organization	134			35	6.41E-13			14	6.60E-04	21	1.17E-05
90304	nucleic acid metabolic process	1510			186	3.45E-32			84	1.31E-15	91	4.20E-04

Table 4-8: Functional class enrichment analysis of genes up-regulated by mistranslation in *S. cerevisiae* grown at 30ºC. The Gene Ontology enrichment analysis was done using genes with a fold change >+1.5 and <-1.5. (Complete table in annexes, table 16).

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
5991	trehalose metabolic process	11			8	7.76E-05			6	3.17E-03	8	3.20E-06
15980	energy derivation by oxidation of organic compounds	138			40	5.87E-12			28	4.77E-08	27	3.49E-07
16052	carbohydrate catabolic process	103			23	6.80E-04	9	3.08E-03	17	5.62E-03	20	7.38E-05
42775	mitochondrial ATP synthesis coupled electron transport	28			12	1.50E-04			11	2.15E-05	11	2.51E-05
55114	oxidation-reduction process	171			46	1.19E-12	13	1.60E-04	34	7.25E-10	34	1.10E-09

4.4.3. Common response to mistranslation

To elucidate whether the transcriptome profiles were *C. albicans* and *S. cerevisiae* specific or conditions (30º or 37ºC) specific, rather than associated with the chemical properties of the amino acids replaced by Ser misincorporation, the list of deregulated genes were compared. Since *C. albicans* deregulated fewer genes than *S. cerevisiae* it was assumed that a common response only takes place in *C. albicans* if genes were deregulated at least in three strains. For *S. cerevisiae* it was assumed that a common response involved deregulated genes in at least four strains.

Mistranslating *C. albicans* grown at 30ºC

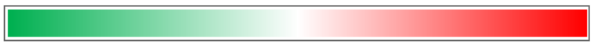
The majority of shared deregulated genes are linked to transport. In particular *orf19.1867*, *orf19.2848*, *FCY24* and *YHB4* show higher deregulation level (Table 4-9). *orf19.1867* encodes a putative peroxisomal protein, namely the malate permease (Mae1p), which likely plays a key role in malate transport across the peroxisome membrane. Its expression is upregulated during phagocytosis, which suggests that *C. albicans* requires malate-dependent metabolism during this process (Prigneau *et al.*, 2003). The *orf19.2848* is not yet characterized in *C. albicans*, however its *S. cerevisiae* orthologous *ATG13* encodes a protein that regulates the autophagosome formation. This process occurs through Atg13 binding to protein Atg1, leading to its activation and induction of the mechanisms of autophagy (Suzuki and Ohsumi, 2007).

Fcy24 is a putative transporter of vitamin B6 controlled by Nrg1, which is responsible for regulating various cellular functions, including some stress responses, morphogenesis and virulence (Murad *et al.*, 2001). YHB4 is responsible for a predicted flavohemoglobin with FAD-binding, and NAD(P)-binding domains, but lacking the conserved Tyr B10 in the heme pocket, which is essential for high oxygen (O_2) affinity and efficient nitric oxide (NO) dioxygenation. yhb4Δ null mutants do not show nitric oxide (NO) consumption decrease (Ullmann *et al.*, 2004).

RAD10 belongs to DNA metabolic processes, is shared among four of the studied strains, namely strains misincorporating Ser at Ala, Leu, Lys and Tyr codons. Although this gene is not characterized in *C. albicans*, the orthologous *S. cerevisiae* gene encodes a DNA endonuclease involved in DNA damage recognition and nucleotide excision repair (Prakash and Prakash, 2000).This suggests that mistranslation may induce DNA damage and may even lead to genome instability. Interestingly, RNA metabolic processes are also affected since *orf19.265* showed high fold variation. The function of this gene is not yet fully understood, but it may also be linked to biofilm formation (Bonhomme *et al.*, 2011). The GO term response to chemical stimulus identified *IME2* has the most deregulated gene of this functional category. The *IME2* functions has not yet been

identified in *C. albicans*, but its *S. cerevisiae* ortholog is a Ser/Thr protein kinase involved in the regulation of meiosis and sporulation (Guttmann-Raviv *et al.*, 2002). Among the pathogenesis genes, *IFF11* was the most deregulated and encodes a protein that belongs to the Iff family (cell wall-related proteins). Iff11 lacks the GPI-anchor and a transmembrane domain, and is secreted, playing a role in virulence and cell wall stress. In fact, a iff11Δ null mutant exhibited attenuated infection capacity and hypersensitivity when exposed to cell wall perturbing agents like calcofluor white, Congo red and SDS (Bates *et al.*, 2007).

Table 4-9: Transcriptional response of *C. albicans* misincorporating Ser at Ala, Leu, Lys, Thr or Tyr codons, grown at 30ºC. Table shows overlapping of deregulated genes (Complete table in annexes table 17).

-3  3

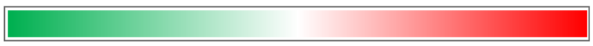
Process	Gene Id	Fold				
		Ala	Leu	Lys	Thr	Tyr
cellular protein modification process	CWH43	-1.22	-1.36			-1.14
	orf19.3267	1.22	-1.25			-1.29
DNA metabolic process	RAD10	1.75	1.43	1.28		1.29
pathogenesis	IFF11	1.63	8.34		1.78	
response to chemical stimulus /drug	IME2	5.28	2.35			1.64
	orf19.5621	-2.04	-2.06			-1.88
RNA metabolic process	orf19.265	222.56	8.65		74.54	
	orf19.332	-1.36	1.54			1.44
	SMD2	-1.47	-2.38			-1.53
	RPL35	-1.32	-1.43			-1.68
transport	orf19.1142	1.55	1.49		1.39	
	FCY21	1.41	1.73			1.35
	orf19.1867	20.19	6.13		12.15	
	orf19.2848	3.41	1.70			1.70
	APM1	-1.58	-1.34			-1.51
	PHO89	1.49	1.63			1.39
	orf19.6435	1.24	-1.35			-1.38
	DUR7	1.26	1.36		1.26	
	FCY24	2.65	3.02			2.20
	YHB4	2.31	2.74	1.31		
vitamin metabolic process	THI6	1.67	1.46			1.38

Mistranslating *C. albicans* grown at 37ºC

The majority of shared genes deregulated by *C. albicans* grown at 37ºC were repressed and most of them are implicated in transport processes (Table 4-10). However, other processes were also affected. For example, *orf19.4581*

belongs to cell wall organization process and is linked to the response to hydrogen peroxide (H_2O_2) was highly deregulated (Urban *et al.*, 2005). The *HWP2* gene was also deregulated across strains. The hwp2 protein is involved in interaction between organisms namely with human cells and encodes a cell wall GPI-anchored protein necessary for hyphal and invasive growth in solid medium (Hayek *et al.*, 2010). It is also involved in tolerance to H_2O_2, since hwp2Δ null mutant is more susceptible to H_2O_2 (Younes *et al.*, 2011). The *STP2*, *QDR1* and *FCY24* genes are involved in transport processes were deregulated across strains. Stp2 is a transcription factor responsible for regulating expression of amino acid permeases, which facilitate amino acid uptake (Younes *et al.*, 2011). Qdr1 and Fcy24 are putative transporters regulated by Nrg1, which regulates various cellular functions, including some stress responses, morphogenesis and virulence (Murad *et al.*, 2001).

Table 4-10: Overlap transcriptional responses of *C. albicans* misincorporating Ser at Ala, Leu, Lys, Thr and Tyr codons, grown at 37ºC. (Complete table in annexes table 18).

-3  3

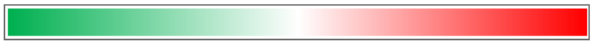
		Fold				
Process	**Gene Id**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
cell wall organization	orf19.4581	-1.71		-1.69	-2.12	
interspecies interaction between organisms	HWP2		-1.38	-2.37	-2.16	
protein catabolic process	orf19.6630	1.38		-1.29		1.39
regulation of biological process	ZCF22	-1.58		-1.32		-1.81
	orf19.6810	-1.29	-1.38			-1.30
RNA metabolic process	PRP22	1.10	1.14			1.09
transport	IFC1	-2.32			-1.77	1.37
	orf19.4228	-1.11		-1.09		-1.06
	orf19.4811	1.23		-1.41		1.35
	STP2	-1.48	-1.63			-1.47
	QDR1			-1.51	-1.71	-1.51
	FCY24	-1.75	-1.62	-1.43		
vitamin metabolic process	THI20		-1.39	-1.21	-1.29	1.37

Mistranslating *S. cerevisiae* strains

In *S. cerevisiae* down regulated genes belong mainly to processes related to protein synthesis namely, cytoplasmatic translation, ribosomal large subunit biogenesis, peptidyl-amino acid modification and protein maturation. The up-

regulated genes are involved in cell wall organization and biogenesis, cellular amino acid metabolic process, cofactor metabolic process, histone modification/chromatin processes, lipid metabolic processes, membrane fusion, mitochondria organization, protein complex biogenesis and phosphorylation, response to chemical stimulus, sporulation, RNA polymerase II transcription and transmembrane transport.

Table 4-11: Overlap transcriptional responses of *S. cerevisiae* misincorporating Ser at Ala, Leu, Lys, Thr and Tyr codons, grown at 30ºC. (Complete table in annexes table 19).

-3  3

		Fold				
Process	**Gene Id**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
cell wall organization or biogenesis	RCR1	1.3	1.3		1.3	1.4
cellular amino acid metabolic process	AGX1	1.7	4.2	1.7	2.3	2.9
	ALT1	1.3	1.7		1.3	1.6
cofactor metabolic process	YNL200c		2.4	1.4	1.7	1.6
	COQ9		1.5	1.3	1.5	1.4
cytoplasmic translation	TMA46		-1.7	-1.3	-1.9	-1.5
	RBG2	-1.3	-1.7		-1.6	-1.5
histone modification/ chromatin organization	JHD2		1.4	1.3	1.4	1.5
lipid metabolic process	CLD1	2.1	2.8	1.5	1.8	2.6
	PGC1	1.3	1.4	1.5	1.6	1.5
	UPS1	1.1	1.2	1.2		1.1
	GDE1	-1.1	1.6	-1.2	1.3	
membrane fusion	YHR138c	1.5	2.3		1.6	1.8
mitochondrion organization	GEM1	1.2	1.3		1.3	1.2
	YDR379c-a	1.5	2.8	1.6	2.6	2.5
peptidyl-amino acid modification	SEE1	-1.3	-2.3		-1.6	-1.6
	YBR271w	-1.3	-1.9		-1.7	-1.5
	HIF1		-1.2	1.1	-1.2	-1.2
protein complex biogenesis	PBA1	1.2	1.4	1.2	1.2	1.3
protein maturation	ICP55	-1.2		-1.1	-1.2	-1.1
protein phosphorylation	YPL150w	1.1	1.3		1.1	1.2
proteolysis	YDR161w	-1.4	-1.9		-1.9	-1.7
response to chemical stimulus	YHI9	1.6	1.9	1.2		1.8
ribosomal large subunit biogenesis	RRP14		-2.1	-2.4	-2.2	-2.0
	URB1	-1.2	-2.0	-1.2	-1.5	-1.6
RNA modification/ tRNA processing	TRM13	-1.3	-2.4	-1.2	-1.7	-1.3
sporulation	YNL194c	2.3	8.9	2.7	4.1	3.7
	FMP45		4.3	1.7	2.6	2.7
transcription from RNA polymerase II promoter	YOR338w	2.1	2.1	1.6	1.9	2.0
	FRA1	1.3	1.4		1.3	1.4
	YRM1	1.2	1.3		1.1	1.2
	CRF1	1.4	2.1		1.4	1.8
transmembrane transport	YBR241c	1.4	2.4	1.6	1.7	1.9
	YDL199c	1.6	2.2		1.7	2.2
	QDR3	1.3	1.4		1.2	1.4

The *AGX1*, *CLD1*, *YDR379c-a*,*YNL194c* and *FMP45* genes were the most deregulated. The *AGX1* gene encodes an alanine:glyoxylate aminotransferase, an enzyme responsible for glycine synthesis (Schlosser *et al.*, 2004). *CLD1* encodes a mitochondrial cardiolipin-specific phospholipase that helps in the maturation of cardiolipin (a dimeric phospholipid abundant in mitochondrial membranes involved in mitochondrial structure and stabilizing respiratory chain super complexes and individual electron transport complexes) (Beranek *et al.*, 2009). Previous studies have shown that this gene is induced in cells exposed to genotoxic stress (Caba *et al.*, 2005). The gene *YDR379c-a* (also named *SDH6*) is involved in the assembly of the succinate dehydrogenase complex in the mitochondria. The other gene that was highly deregulated was *YNL194c*, which is normally up-regulated in response to the DNA-damaging agent, methyl methanesulphonate (Lee *et al.*, 2007).

4.4.4. Transcription factors

Previous studies, using inducible mistranslation in *S. cerevisiae* showed that the main transcription factors motifs present in the deregulated genes are the general stress response element (STRE; AGGGGA/T), the heat-shock responsive element (HSE; nGAAn), the pleiotropic drug resistance element (PDRE; TCCGCGGA targeted by Pdr1p/Pdr3p) and the proteasome associated control element (PACE; GGTGGCAAA; targeted by Rpn4p) (Paredes *et al.*, 2012). In *C. albicans* Enjalbert and coworkers described Hog1 as a transcription factor involved in the core stress response more precisely in osmotic and heavy metal stresses (Enjalbert *et al.*, 2006). The Hog1 transcription factor induces 24 genes involved in cell wall biogenesis, carbohydrate metabolism, drug resistance, mitochondrial respiratory function, and represses 37 genes that are mainly involved in protein synthesis and RNA processing (Enjalbert *et al.*, 2006).

To identify the transcription factors that regulate the response to mistranslation in *C. albicans* a sequence of 1000 base pairs upstream of each gene open reading frame start codon (AUG) was submitted to the motive finder “Find TF Binding Site(s)” in http://www.yeastract.com (Monteiro *et al.*, 2008). The regular transcription factor motifs occurrence was established using 100 random genes, which was used for comparative purposes. In our deregulated genes

dataset, several transcription factor motifs show a frequency of occurrence that was different from that obtained with the random gene motif list.

S. cerevisiae showed an enrichment of motifs that are recognized by Gis1, Pho4 and Abf1 transcription factors (Table 4-12). The GIS1 transcription factor is associated with the activation of genes that respond to starvation of glucose, ammonium and phosphate (Zhang *et al.*, 2009). One mechanism that controls the cellular abundance of Gis1p is its degradation by the proteasome (Zhang and Oliver, 2010), suggesting that the proteasome may be saturated by mistranslated proteins, resulting in a lower degradation rate of Gis1p. Subsequently, the accumulation of this protein leads to expression of genes related to nutrient starvation. The Pho4 transcription factor is involved in the response to inorganic phosphate availability. When cells are limited in inorganic phosphate, Pho4 is dephosphorylated and transported from the cytoplasm to the nucleus, activating a transcriptional response involving the transcription factor Pho2. On the other hand, when cells have inorganic phosphate, Pho4 is phosphorylated and transported from the nucleus to the cytoplasm (Zhou and O'Shea, 2011). The Abf1 transcription factor is less frequent in our deregulated gene dataset. This transcription factor is responsible for organizing chromatin structure (Lascaris *et al.*, 2000), and is involved in DNA replication, transcription, gene silencing, and is also directly involved in nucleotide excision repair (Reed *et al.*, 1999). Because Abf1 associates with the Rad7/Rad16 protein complex of the nucleotide excision repair process (Reed *et al.*, 1999), and since Rad16 gene is upregulated in 3 (Alanine=1.367493, Leucine=1.442676, Tyrosine=1.575551) of the 5 studied strains, it is likely that the Abf1 is captured by the Rad7/Rad6 complex resulting in lower Abf1 concentration and subsequently in a lower deregulation of genes controlled by this transcription factor.

Table 4-12: Transcription factors that recognized motifs in the deregulated gene dataset from *C. albicans* and *S. cerevisiae*. (*) motifs that are recognized by a specific transcription factor is *S. cerevisiae* but there is no described orthologous gene in *C. albicans* (complete table in annexes, table 20).

Transcription factor		*C. albicans* (%)			*S.cerevisiae* (%)	
S.cerevisiae	*C.albicans* orthologs Gene Name	30ºC	37ºC	Random	30ºC	Random
Abf1p	*	0.0	0.0	6.0	46.1	60.2
Gat1p	GAT1 (ORF19.1275)	48.7	78.9	30.0	45.3	40.8
Gis1p	*	51.3	52.6	57.0	68.8	54.1
Gln3p	GLN3(ORF19.3912)	48.7	78.9	30.0	57.8	58.2
Gzf3p	GZF3	48.7	78.9	30.0	45.3	40.8
Hap2p	HAP2	46.2	42.1	33.0	14.1	11.2
Hap3p	HAP3 (ORF19.4647)	46.2	42.1	33.0	14.1	11.2
Hap4p	HAP43	46.2	42.1	33.0	14.1	11.2
Hap5p	HAP5	46.2	42.1	33.0	14.1	11.2
Pho4p	ORF19.1253	30.8	26.3	33.0	41.4	30.6

In *C. albicans*, the transcription factor motifs that showed higher enrichment in the deregulated overlapped genes (complete list of genes from Table 4-9 and Table 4-10) were Gat1, Gln3, and Gzf3 (Table 4-12). Gat1 and Gln3 are GATA transcription factors that control the expression of the ammonium permease *MEP2*. In addition, Gln3 is also a regulator of nitrogen starvation-induced filamentous growth (Dabas and Morschhauser, 2007).

GZF3 is the glucose-6-phosphate dehydrogenase gene and it is also a putative transcription factor induced by Cap1 when cells are exposed to oxidative stress. It is often suggested that this pathway plays a role in the oxidative stress response through the main NADPH-producing pathway, namely the pentose phosphate pathway (Wang *et al.*, 2006). Another transcription factor motif that was frequent in the list of deregulated genes was the Stb5 motif. Stb5 is a transcription factor that is repressed by Hap43 (Singh *et al.*, 2011). Although it is not fully characterized in *C. albicans* it shares high level of identity with the *S. cerevisiae* homologue (Maicas *et al.*, 2005), being responsible in yeast for regulating the

expression of genes belonging to the pentose phosphate pathway, as well as, genes responsible in for NADPH production, a metabolite involved in oxidative stress response (Larochelle *et al.*, 2006).

Another set of transcription factor motifs that appeared frequently in the overlapped deregulated gene list were the transcription factor Hap2, Hap3, Hap4 and Hap5. These transcription factors and Hap43 are involved in the iron homeostasis. They form a complex that binds to a ferric reductase *FRP1* promoter under iron-deprivation conditions. Subsequently, this ferric reductase converts ferric iron (Fe^{3+}) to the biologically available ferrous iron (Fe^{2+}) and promotes high-affinity iron transport (Baek *et al.*, 2008;Singh *et al.*, 2011). Another functional model is that the Hap complex works as an anti-repressor, destabilizing a putative repressor under low iron conditions permitting assembly of the transcriptional machinery (Singh *et al.*, 2011). The iron is required for several biochemical processes, namely cellular respiration, metabolism and DNA synthesis (Welch *et al.*, 2001). In the case of *C. albicans* iron uptake is needed for proliferation and colonization and it is considered a virulence attribute (Sutak *et al.*, 2008;Almeida *et al.*, 2009).

4.5. Discussion

In this study we have engineered new amino acids misincorporations into *C. albicans* to clarify how this fungus responds to genetic code ambiguity. The constructed strains misincorporate Ser at Leu, Ala, Lys, Thr and Tyr codons. Additionally, *S. cerevisiae* strains misincorporating Ser in the same amino acids were also studied. The purpose of the latter was to compare the cellular responses to mistranslation of a naturally ambiguous organism (*C. albicans*) with those of an organism not exposed to high level codon ambiguity (*S. cerevisiae*).

In S. cerevisiae, the different mistranslating strains deregulated a common group of genes, suggesting the activation of a core stress response. Conversely, in C. *albicans* shared deregulated genes between the different mistranslating strains were not observed, suggesting that this organism does not respond to

mistranslation through activation of a cross stress response. These results are in line with the studies carried by Enjalbert and colleagues, who showed that the stress response in *C. albicans* involves a small subset of stress genes (3.7%) and there is no conserved core stress response (Enjalbert *et al.*, 2006). Since the end points of mistranslation are protein misfolding and aggregation, it is unclear why the *C. albicans* misincorporating strains did not activate a common set of genes. However a likely explanation is that *C. albicans* is pre-adapted to the proteotoxic stress induced by mistranslation due to ambiguous decoding of the CUG codon (Gomes *et al.*, 2007;Rocha *et al.*, 2011b)

The *C. albicans* strains grown at 30ºC repressed genes involved in the regulation of gene expression, in particular the strains that misincorporate Ser at Leu and Tyr codons. Ribosome biogenesis genes were repressed in the strains that misincorporated Ser at Ala and Leu codons. The lower abundance of mRNAs encoding proteins involved in ribosome biogenesis was in line with the lower protein synthesis rate observed in strains misincorporating Ser at Ala codons (chapter 3). Up-regulated genes shared between strains misincorporating Ser at Ala, Leu and Lys codons are involved in oxidation reduction processes. It was also interesting that *C. albicans* misincorporating strains grown at 37ºC, shared almost no repressed genes, the exception being the genes involved transmembrane transport process. However, several up-regulated genes involved in glucose catabolic process, oxidative phosphorylation, mitochondrial electron transport of NADH to ubiquinone, respiratory electron transport chain, generation of precursor and metabolites and energy processes were observed across the strains, suggesting that mitochondrial function was affected in all strains tested. Finally, *S. cerevisiae* showed higher number of shared deregulated genes and processes among the different mistranslating strains. The majority of repressed genes are involved in ribosomal biogenesis and assembly processes, in particular in strains misincorporating Ser at Leu, Thr and Tyr codons. Previous work using inducible mistranslation in *S. cerevisiae* revealed similar repression of ribosomal biogenesis and assembly processes (Paredes *et al.*, 2012). Regarding the up-regulated genes, the higher number of affected genes belong to the glycogen and trehalose metabolic processes (strains mistranslating Leu, Thr and Tyr codons) and

oxidation reduction processes (strains mistranslating Leu, Lys, Thr and Tyr codons). Previous work using constitutive and inducible mistranslation in *S. cerevisiae* showed identical results (Silva *et al.*, 2004;Silva *et al.*, 2007;Paredes *et al.*, 2012). Indeed, trehalose stabilizes proteins in yeast cells during heat shock and suppresses the aggregation of denatured proteins (Singer and Lindquist, 1998). The up regulation of the oxidation reduction process observed in *S. cerevisiae* and *C. albicans* grown at 30ºC, the deregulation of the mitochondrial electron transport of NADH to ubiquinone, the respiratory electron transport chain, the generation of precursor and metabolites and the energy process in *C. albicans* grown at 37ºC, strongly suggest therefore that mitochondria are major targets of mistranslation.

Venn diagrams of deregulated genes show that mistranslation does not deregulate the same genes in *S. cerevisiae* and *C. albicans* misincorporating Ser at identical amino acids. This provides strong evidence for differential responses to mistranslation between these two yeasts. Besides the small number of homologous genes deregulated between *S. cerevisiae* and *C. albicans* misincorporating strains, there were also few shared deregulated genes between *C*. *albicans* strains grown at 30ºC and 37ºC, suggesting that the response to mistranslation may also be affected by environmental cues. Interestingly previous studies have also shown that, unlike *S. cerevisiae*, *C. albicans* exposed to oxidative, osmotic or heavy metal does not activate a common stress response (Enjalbert *et al.*, 2006), suggesting that it responds to each stress in a specific manner and lacks stress cross-protection mechanisms. This raises the possibility that each amino acid misincorporation is sensed as a unique stress by *C. albicans* cells. In other words, the stress responses in *C. albicans* and *S. cerevisiae* are fundamentally different.

In line with the above hypothesis is our finding that the transcription factors that regulate the response to mistranslation are also different between *C. albicans* and *S. cerevisiae* (Paredes *et al.*, 2012). The transcription factors motifs that occurred at most of the deregulated genes were Gis1 and Pho4. Gis1 is involved in the activation of a set of genes that are expressed in response to glucose, ammonium and phosphate starvation (Zhang *et al.*, 2009). The Pho4 transcription

factor regulates the response to inorganic phosphate availability. When cells have limited levels of inorganic phosphate, Pho4 interacts with the transcription factor Pho2 to activate the expression of various genes involved in phosphate metabolism (Zhou and O'Shea, 2011). The apparent repression of the Abf1 transcription factor was also interesting since this transcription factor is responsible for expression of genes that organize chromatin structure (Lascaris *et al.*, 2000), and is implicated in DNA replication, transcription, gene silencing, and nucleotide excision repair (Reed *et al.*, 1999). This suggests that our *S. cerevisiae* strains may have higher rate of mutagenesis than the WT strains.

The *C. albicans* transcription factors motifs that occurred more frequently in the deregulated shared genes were Gat1, Gln3, and Gzf3. Gat1 and Gln3 are GATA transcription factors that control the expression of the ammonium permease *MEP2* gene. The Gln3 is a regulator of nitrogen starvation-induced filamentous growth (Dabas and Morschhauser, 2007). Gzf3 is a putative transcription factor induced via Cap1 when cells are exposed to oxidative stress (Wang *et al.*, 2006). This pathway could play roles in the oxidative stress response through the main NADPH-producing pathway, namely the pentose phosphate pathway (Wang *et al.*, 2006). Another transcription factor motif whose *cis*-element was less frequent in the deregulated genes was Stb5, although it is not fully characterized in *C. albicans*. The *S. cerevisiae* homologue is implicated in regulation of genes belonging to the pentose phosphate pathway and NADPH production, a metabolite implicated in oxidative stress response (Larochelle *et al.*, 2006). Finally the cis-element recognized by the Hap2, Hap3, Hap4 and Hap5 transcription factors were also over-represented in the deregulated genes list. These transcription factors work in coordination with Hap43 in the control of iron homeostasis (Baek *et al.*, 2008;Singh *et al.*, 2011). Iron is essential for several processes including DNA synthesis and cellular respiration (Welch *et al.*, 2001). Furthermore, the iron uptake in *C. albicans* is a requirement for virulence since it permits proliferation and colonization of the host (Sutak *et al.*, 2008;Almeida *et al.*, 2009). This observation opens the possibility of mistranslation working as a catalyst to colonization of new environmental niches and even virulence. Overall, the transcription factors enrichment analysis showed that misincorporation of Ser at

amino acid codons is a major cause of ROS, which is in agreement with previous studies carried out by Paredes (Paredes *et al.*, 2012).

4.6. Conclusion

Misincorporation of Ser at non-cognate codons deregulated higher number of genes and also higher number of overlapping genes in *S. cerevisiae* than in *C. albicans*. Interestingly, there was no common response to mistranslation between *C. albicans* and *S. cerevisiae*, indicating that both species sense and respond to protein aggregation in fundamentally different ways. The reasons for these surprising differences are not clear, but will be fascinating to evolve tolerance to Ser misincorporation in *S. cerevisiae* and determine whether the resulting evolved strains reprogram the stress response during the course of such evolution in similar ways to *C. albicans*.

4.7. References

Almeida,R.S., Wilson,D., and Hube,B. (2009). *Candida albicans* iron acquisition within the host. FEMS Yeast Res. *9*, 1000-1012.

Ambrogelly,A., Palioura,S., and Soll,D. (2007). Natural expansion of the genetic code. Nat. Chem. Biol. *3*, 29-35.

Baek,Y.U., Li,M., and Davis,D.A. (2008). *Candida albicans* ferric reductases are differentially regulated in response to distinct forms of iron limitation by the Rim101 and CBF transcription factors. Eukaryot. Cell *7*, 1168-1179.

Bates,S., de la Rosa,J.M., MacCallum,D.M., Brown,A.J., Gow,N.A., and Odds,F.C. (2007). *Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. Infect. Immun. *75*, 2922-2928.

Beranek,A., Rechberger,G., Knauer,H., Wolinski,H., Kohlwein,S.D., and Leber,R. (2009). Identification of a cardiolipin-specific phospholipase encoded by the gene CLD1 (YGR110W) in yeast. J. Biol. Chem. *284*, 11572-11578.

Bonhomme,J., Chauvel,M., Goyard,S., Roux,P., Rossignol,T., and d'Enfert,C. (2011). Contribution of the glycolytic flux and hypoxia adaptation to efficient biofilm formation by *Candida albicans*. Mol. Microbiol. *80*, 995-1013.

Caba,E., Dickinson,D.A., Warnes,G.R., and Aubrecht,J. (2005). Differentiating mechanisms of toxicity using global gene expression analysis in *Saccharomyces cerevisiae*. Mutat. Res. *575*, 34-46.

Causton,H.C., Ren,B., Koh,S.S., Harbison,C.T., Kanin,E., Jennings,E.G., Lee,T.I., True,H.L., Lander,E.S., and Young,R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell *12*, 323-337.

Dabas,N. and Morschhauser,J. (2007). Control of ammonium permease expression and filamentous growth by the GATA transcription factors GLN3 and GAT1 in *Candida albicans*. Eukaryot. Cell *6*, 875-888.

Di,G.M. (2005). The origin of the genetic code: theories and their relationships, a review. Biosystems *80*, 175-184.

Drummond,D.A. and Wilke,C.O. (2009). The evolutionary consequences of erroneous protein synthesis. Nat. Rev. Genet. *10*, 715-724.

Enjalbert,B., Nantel,A., and Whiteway,M. (2003). Stress-induced gene expression in Candida albicans: absence of a general stress response. Mol. Biol. Cell *14*, 1460-1467.

Enjalbert,B., Smith,D.A., Cornell,M.J., Alam,I., Nicholls,S., Brown,A.J., and Quinn,J. (2006). Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen *Candida albicans*. Mol. Biol. Cell *17*, 1018-1032.

Gasch,A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D., and Brown,P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241-4257.

Gasch,A.P. and Werner-Washburne,M. (2002). The genomics of yeast responses to environmental stress and starvation. Funct. Integr. Genomics *2*, 181-192.

Gentleman,R.C. *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206.

Guttmann-Raviv,N., Martin,S., and Kassir,Y. (2002). Ime2, a meiosis-specific kinase in yeast, is required for destabilization of its transcriptional activator, Ime1. Mol. Cell Biol. *22*, 2047-2056.

Harnpicharnchai,P. *et al.* (2001). Composition and functional characterization of yeast 66S ribosome assembly intermediates. Mol. Cell *8*, 505-515.

Hayek,P., Dib,L., Yazbeck,P., Beyrouthy,B., and Khalaf,R.A. (2010). Characterization of Hwp2, a *Candida albicans* putative GPI-anchored cell wall protein necessary for invasive growth. Microbiol. Res. *165*, 250-258.

Homann,O.R., Dea,J., Noble,S.M., and Johnson,A.D. (2009). A phenotypic profile of the *Candida albicans* regulatory network. PLoS. Genet. *5*, e1000783.

Kitanovic,A. and Wolfl,S. (2006). Fructose-1,6-bisphosphatase mediates cellular responses to DNA damage and aging in *Saccharomyces cerevisiae*. Mutat. Res. *594*, 135-147.

Knight,R.D., Freeland,S.J., and Landweber,L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet *2*, 49-58.

Larochelle,M., Drouin,S., Robert,F., and Turcotte,B. (2006). Oxidative stress-activated zinc cluster protein Stb5 has dual activator/repressor functions required for pentose phosphate pathway regulation and NADPH production. Mol. Cell Biol. *26*, 6690-6701.

Lascaris,R.F., Groot,E., Hoen,P.B., Mager,W.H., and Planta,R.J. (2000). Different roles for abf1p and a T-rich promoter element in nucleosome organization of the yeast RPS28A gene. Nucleic Acids Res. *28*, 1390-1396.

Lee,M.W., Kim,B.J., Choi,H.K., Ryu,M.J., Kim,S.B., Kang,K.M., Cho,E.J., Youn,H.D., Huh,W.K., and Kim,S.T. (2007). Global protein expression profiling of budding yeast in response to DNA damage. Yeast *24*, 145-154.

Lenhard,B., Orellana,O., Ibba,M., and Weygand-Durasevic,I. (1999). tRNA recognition and evolution of determinants in seryl-tRNA synthesis. Nucleic Acids Res. *27*, 721-729.

Maicas,S., Moreno,I., Nieto,A., Gomez,M., Sentandreu,R., and Valentin,E. (2005). In silico analysis for transcription factors with Zn(II)(2)C(6) binuclear cluster DNA-binding domains in Candida albicans. Comp Funct. Genomics *6*, 345-356.

Massey,S.E., Moura,G., Beltrao,P., Almeida,R., Garey,J.R., Tuite,M.F., and Santos,M.A. (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. Genome Res. *13*, 544-557.

Mateus,D.D. Molecular reconstruction of a genetic code alteration. 6-17-2011. Universidade de Aveiro. Ref Type: Thesis/Dissertation

Mateus,D.D., Paredes,J.A., Espanol,Y., Ribas de,P.L., Moura,G.R., and Santos,M.A. (2013). Molecular reconstruction of a fungal genetic code alteration. RNA. Biol. *10*.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A Genetic Code Alteration Is a Phenotype Diversity Generator in the Human Pathogen *Candida albicans*. PLoS. ONE. *2*, e996.

Miranda,I., Silva,R., and Santos,M.A. (2006). Evolution of the genetic code in yeasts. Yeast *23*, 203-213.

Monteiro,P.T. *et al.* (2008). YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. Nucleic Acids Res. *36*, D132-D136.

Murad,A.M., d'Enfert,C., Gaillardin,C., Tournu,H., Tekaia,F., Talibi,D., Marechal,D., Marchais,V., Cottin,J., and Brown,A.J. (2001). Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. Mol. Microbiol. *42*, 981-993.

Nicholls,S., Straffon,M., Enjalbert,B., Nantel,A., Macaskill,S., Whiteway,M., and Brown,A.J. (2004). Msn2- and Msn4-like transcription factors play no obvious roles in the stress responses of the fungal pathogen *Candida albicans*. Eukaryot. Cell *3*, 1111-1123.

Noble,S.M. and Johnson,A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. Eukaryot. Cell *4*, 298-309.

Paredes,J.A., Carreto,L., Simoes,J., Bezerra,A.R., Gomes,A.C., Santamaria,R., Kapushesky,M., Moura,G.R., and Santos,M.A. (2012). Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC. Biol. *10*, 55.

Polevoda,B., Arnesen,T., and Sherman,F. (2009). A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. BMC. Proc. *3 Suppl 6*, S2.

Prakash,S. and Prakash,L. (2000). Nucleotide excision repair in yeast. Mutat. Res. *451*, 13-24.

Prigneau,O., Porta,A., Poudrier,J.A., Colonna-Romano,S., Noel,T., and Maresca,B. (2003). Genes involved in beta-oxidation, energy metabolism and glyoxylate cycle are induced by *Candida albicans* during macrophage infection. Yeast *20*, 723-730.

Reed,S.H., Akiyama,M., Stillman,B., and Friedberg,E.C. (1999). Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. Genes Dev. *13*, 3052-3058.

Rocha,R., Pereira,P.J., Santos,M.A., and edo-Ribeiro,S. (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc. Natl. Acad. Sci. U. S. A *108*, 14091-14096.

Santos,M.A., Cheesman,C., Costa,V., Moradas-Ferreira,P., and Tuite,M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Mol. Microbiol. *31*, 937-947.

Santos,M.A., Gomes,A.C., Santos,M.C., Carreto,L.C., and Moura,G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607-611.

Santos,M.A., Keith,G., and Tuite,M.F. (1993). Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. EMBO J. *12*, 607-616.

Santos,M.A., Perreau,V.M., and Tuite,M.F. (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. EMBO J. *15*, 5060-5068.

Santos,M.A. and Tuite,M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res. *23*, 1481-1486.

Santos,M.A., Ueda,T., Watanabe,K., and Tuite,M.F. (1997). The non-standard genetic code of *Candida spp*.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423-431.

Schlosser,T., Gatgens,C., Weber,U., and Stahmann,K.P. (2004). Alanine : glyoxylate aminotransferase of *Saccharomyces cerevisiae*-encoding gene AGX1 and metabolic significance. Yeast *21*, 63-73.

Schmitt,M.E., Brown,T.A., and Trumpower,B.L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res. *18*, 3091-3092.

Silva,R.M., Miranda,I., Moura,G., and Santos,M.A. (2004). Yeast as a model organism for studying the evolution of non-standard genetic codes. Brief. Funct. Genomic. Proteomic. *3*, 35-46.

Silva,R.M. *et al.* (2007). Critical roles for a genetic code alteration in the evolution of the genus *Candida*. EMBO J. *26*, 4555-4565.

Singer,M.A. and Lindquist,S. (1998). Multiple Effects of Trehalose on Protein Folding In Vitro and In Vivo. Molecular Cell *1*, 639-648.

Singh,R.P., Prasad,H.K., Sinha,I., Agarwal,N., and Natarajan,K. (2011). Cap2-HAP complex is a critical transcriptional regulator that has dual but contrasting roles in regulation of iron homeostasis in *Candida albicans*. J. Biol. Chem. *286*, 25154-25170.

Soll,D. and RajBhandary,U.L. (2006). The genetic code - thawing the 'frozen accident'. J. Biosci. *31*, 459-463.

Srinivasan,G., James,C.M., and Krzycki,J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Sutak,R., Lesuisse,E., Tachezy,J., and Richardson,D.R. (2008). Crusade for iron: iron uptake in unicellular eukaryotes and its significance for virulence. Trends Microbiol. *16*, 261-268.

Suzuki,K. and Ohsumi,Y. (2007). Molecular machinery of autophagosome formation in yeast, *Saccharomyces cerevisiae*. FEBS Letters *581*, 2156-2161.

Suzuki,T., Ueda,T., and Watanabe,K. (1997). The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122-1134.

Ullmann,B.D., Myers,H., Chiranand,W., Lazzell,A.L., Zhao,Q., Vega,L.A., Lopez-Ribot,J.L., Gardner,P.R., and Gustin,M.C. (2004). Inducible defense mechanism against nitric oxide in *Candida albicans*. Eukaryot. Cell *3*, 715-723.

Urban,C., Xiong,X., Sohn,K., Schroppel,K., Brunner,H., and Rupp,S. (2005). The moonlighting protein Tsa1p is implicated in oxidative stress response and in cell wall biogenesis in *Candida albicans*. Mol. Microbiol. *57*, 1318-1341.

Vukotic,M. *et al.* (2012). Rcf1 mediates cytochrome oxidase assembly and respirasome formation, revealing heterogeneity of the enzyme complex. Cell Metab *15*, 336-347.

Walther,A. and Wendland,J. (2003). An improved transformation protocol for the human fungal pathogen *Candida albicans*. Curr. Genet *42*, 339-343.

Wang,Y., Cao,Y.Y., Jia,X.M., Cao,Y.B., Gao,P.H., Fu,X.P., Ying,K., Chen,W.S., and Jiang,Y.Y. (2006). Cap1p is involved in multiple pathways of oxidative stress response in *Candida albicans*. Free Radical Biology and Medicine *40*, 1201-1209.

Welch,K.D., Van Eden,M.E., and Aust,S.D. (2001). Modification of ferritin during iron loading. Free Radic. Biol. Med. *31*, 999-1006.

Younes,S., Bahnan,W., Dimassi,H.I., and Khalaf,R.A. (2011). The *Candida albicans* Hwp2 is necessary for proper adhesion, biofilm formation and oxidative stress tolerance. Microbiol. Res. *166*, 430-436.

Zhang,N. and Oliver,S.G. (2010). The transcription activity of Gis1 is negatively modulated by proteasome-mediated limited proteolysis. J. Biol. Chem. *285*, 6465-6476.

Zhang,N., Wu,J., and Oliver,S.G. (2009). Gis1 is required for transcriptional reprogramming of carbon metabolism and the stress response during transition into stationary phase in yeast. Microbiology *155*, 1690-1698.

Zhou,X. and O'Shea,E.K. (2011). Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. Mol. Cell *42*, 826-836.

5. General discussion

The genetic code is a chemical algorithm that translates the genetic of information stored in DNA in the form of triplets (codons) into proteins formed by joining amino acids blocks (Crick, 1968). It was supposed to be universal and immutable, since any alteration to its rules would be detrimental or lethal to any organism (Crick, 1968). However, the discovery of several codon reassignments namely the stop codons UAA, UAG and UGA and the sense codons AUA, AAA, AGA, AGG, CUG and the four-codon boxes CUN and CGN (Knight *et al.*, 2001;Di, 2005), invalidate the hypothesis of a frozen and universal code. The reports of genetic code alterations started with the discovery of translation of the UGA stop codon as tryptophan and the AUA-isoleucine codon as methionine in the human mitochondria (Barrell *et al.*, 1979). Since this change happened in a relatively small genome it was assumed that the small size of the mitochondria encoded proteome would allow for the tolerance to these genetic code alterations (Ohama *et al.*, 2008). However the discovery of genetic code alterations in *Mycoplasma capricolum*, namely the use of UGA as a tryptophan rather than a stop codon (Yamao *et al.*, 1985) provided evidence for genetic code unanticipated flexibility.

It is now clear that genetic code alterations are widespread in mitochondria. For example, the stop codon UGA is reassigned to tryptophan in the mitochondria of a large number of phyla, the other stop codon UAA is assigned to tyrosine in several *Platyhelminth*, while in the mitochondria of fungi and several plants the stop codon UAG is assigned to alanine and leucine. In *Echinoderm* and *Platyhelminth* mitochondria the lysine AAA codon is assigned to arginine, while in various vertebrates mitochondria the arginine codons AGA and AGG are used as stop codons and in various animal mitochondria as serine codons. Furthermore in the mitochondria of the green algae *Scenedesmus obliques* the serine UCA codon is also a termination codon (Knight *et al.*, 2001;Soll and RajBhandary, 2006).

The UAA, UAG and UGA stop codons have been reassigned in numerous prokaryotic and nuclear genetic codes, in the green algae, ciliates and *Diplomonads* the nuclear codons UAA and UAG encode glutamine, in *Mycoplasma* species, *Spiroplasma citri, Bacillus* and some ciliates the UGA codon is translated as tryptophan and as cysteine (Soll and RajBhandary, 2006). A remarkable nuclear genetic code alteration involves the reassignment of Leu CUG

to Ser in *Candida* species and in many *Ascomycetes* (Santos *et al.*, 1993;Suzuki *et al.*, 1997).

Two naturally occurring amino acids have also been added to the canonical 20 amino acids, namely selenocysteine, which is found in various species from the 3 kingdoms of life and is incorporated in response to programed UGA codon (Ambrogelly *et al.*, 2007) and pyrrolysine, which is incorporated at programed UAG codon in *Methanosarcinaceae* (Srinivasan *et al.*, 2002;Soll and RajBhandary, 2006).

In this thesis we have studied the reassignment of the CUG codon from Leu to Ser in the human pathogen *C. albicans*. This codon is translated ambiguously in various fungal species of the CTG clade because the corresponding tRNA ($tRNA_{CAG}^{Ser}$) is recognized and aminoacylated by both the SerRS and the LeuRS (Suzuki *et al.*, 1997;Gomes *et al.*, 2007). Since Leu-$tRNA_{CAG}^{Ser}$ is not edited by the LeuRS nor discriminated by the translation elongation factor 1 (eEF1A) (Santos *et al.*, 2011), the Leu-$tRNA_{CAG}^{Ser}$ and Ser-$tRNA_{CAG}^{Ser}$ compete for CUG codons during translation incorporating both Leu and Ser at CUG positions (Gomes *et al.*, 2007). Previous studies have demonstrated that when grown in standard laboratory conditions 3% of leucine and 97% of serine are incorporated at CUGs, but when exposed to stress the incorporation of Leu fluctuates between 0.6% and 5% (Gomes *et al.*, 2007).

In this study, we have extended previous analysis and we show that CUG ambiguity fluctuation in response to environmental variation is much higher than previously shown. Our data excludes the hypothesis that *C. albicans* cells regulate CUG ambiguity through a negative feedback mechanism and we show that the *C. albicans* proteome is highly tolerant to genetic code ambiguity.

5.1. CUG ambiguity fluctuates in response to environmental variation

Previous studies have shown that incorporation of Leu at CUG codons is variable, but the scope of such variation is not clearly understood (Gomes *et al.*, 2007). Our new fluorescent reporter system allowed us to obtain a more complete

picture of CUG ambiguity in *C. albicans*. Our data showed very high levels of Leu incorporation at CUGs in presence of macrophages and amphotericin B, suggesting that ambiguity may be relevant to phagocytosis in presence of amphotericin B. These data are in line with previous studies carried out by Miranda and colleagues showing that *C. albicans* strains misincorporating 30% of Leu at CUGs produce heterogeneous cell populations that included elongated-ovoid cells, pseudohypha and hypha, increased cell adhesion, flocculation and production of hydrolases (Miranda *et al.*, 2007). However, how macrophages and the combination of macrophages with amphotericin B increase Leu misincorporation remain to be clarified. A proteome analysis of *C. albicans* exposed to macrophages showed a strong decrease in expression of the SerRS, which may explain the increased Leu misincorporation, however this study needs to be repeated in presence of amphotericin B to clarify putative synergistic effects of macrophages and amphotericin in SerRS repression.

It is also interesting that CUG accumulates preferentially in genes linked to pathogenesis and virulence, biofilm formation, mating and morphogenesis (Rocha *et al.*, 2011). In other words, the increase in Leu incorporation at CUG codons under stress conditions suggest that *C. albicans* responds to stress by increasing the pool of ambiguous proteins (statistical proteins) whether or not they are relevant to stress adaptation needs to be clarified in future studies.

5.2. CUG ambiguity levels and tolerance to proteome disruption

The *C. albicans* $tRNA_{CAG}^{Ser}$ has leucylation and serylation identity elements that permit its recognition by the LeuRS and SerRS. The presence of a CUG codon encoding residue 197 of SerRS, located in the dimer interface, in a position where polar amino acids are preferred in homologous enzymes, suggested that Leu-197 could alter the structure of the motif 3 upstream of the C-terminus tail. Considering the occurrence of two SerRS isoforms, in *C. albicans* cells and that the SerRS_Leu197 is more active *in vitro* than the SerRS_Ser197 isoform, a negative feedback mechanism to keep high levels of Ser-$tRNA_{CAG}^{Ser}$ was proposed (Rocha *et al.*, 2011). According to this mechanism, the increase of Leu

incorporation at SerRS CUG_197 would increase the cellular abundance of the Ser-tRNA$_{CAG}^{Ser}$ relatively to Leu-tRNA$_{CAG}^{Ser}$, resulting in an decreased incorporation of Leu at CUG codons site, thus keeping CUG ambiguity under control (Rocha *et al.*, 2011). With the objective to confirm the existence of this putative mechanism *in vivo* in *C. albicans* we have engineered strains that express only one isoform of the SerRS.

Our results show that the SerRS_Ser197 isoform has higher serylation activity than the SerRS_Leu197, and do not support the negative feedback mechanism proposed previously (Rocha *et al.*, 2011). However, we cannot exclude the possibility that Leu incorporation levels are controlled by the expression levels of SerRS and LeuRS. Our observation that Leu levels increase during stress suggest that the SerRS activity or expression is compromised relative to the LeuRS. Indeed, as mentioned before, previous studies have shown that the cellular abundance of the SerRS decreases in presence of macrophages (Fernandez-Arenas *et al.*, 2007), decreasing the relative SerRS/LeuRS ratio. Since the tRNA$_{CAG}^{Ser}$ is recognized by the LeuRS, lower SerRS/LeuRS ratios lead to increased leucylation of the tRNA$_{CAG}^{Ser}$ by the LeuRs.

To understand if *C. albicans* was able to tolerate ambiguity in other codons, we have engineered strains with a chimeric tRNA that incorporate serine at different non-cognate codons, namely CTA, CTC, CTT (leucine), ATC (isoleucine), GCC (alanine), GGA (glycine), AAG (lysine), ACC (threonine) and TAC (tyrosine) codons. These strains grew slower than the control strains, suggesting that these misincorporations produced some level of proteome disruption. However, they grew better than the control strain in presence of diverse nutritional substrates, temperature, pH and antifungals. These results are in line with previous *S. cerevisiae* studies showing that CUG mistranslation produces selective advantages under specific stress conditions, namely high temperature, salts, heavy metals and oxidants (Santos *et al.*, 1999). Other codon ambiguity studies in *E. coli*, showed that it also produces selective advantage in specific conditions, permitting the colonization of new ecological niches (Pezo *et al.*, 2004;Bacher *et al.*, 2005).

The strains misincorporating Ser at Leu codons were strongly affected and produced aggregated proteins. Interestingly, Ser misincorporation in codons used at low level were more deleterious than Ser misincorporation in codons used at higher level. This suggests that mutant ambiguous tRNAs cause higher proteome disruption if they target codons used at low level, and cause deleterious effects on standard growth conditions, but are less adaptive under stress conditions. Another feature of the strains misincorporating Ser at Thr, Tyr and Leu CTC, CTA codons was the alteration of DNA content. Previous studies in *C. albicans* misincorporating increasing levels of Leu at CUG codons, have also shown increased DNA content levels and ploidy (Miranda *et al.*, 2007). Chromosomal rearrangements and aneuploidy are frequent in *C. albicans*, have significant impact on the catabolic pathways and play important roles in survival and adaptation of microorganisms (Rustchenko, 2007;Ahmad *et al.*, 2008;Bouchonville *et al.*, 2009;Selmecki *et al.*, 2010). These data suggest, therefore, that codon ambiguity may play an important role in the stability of the *C. albicans* genome

Rocha and colleagues conducted a deep analysis of 680 *C. albicans* proteins containing at least one CUG-encoded residue. These data showed that the majority of the *C. albicans* CUG codons are located in nonconserved sequence positions where both Leu (aliphatic amino acid) and Ser (polar amino acid) could be inserted without causing protein structural disruption. Conversely in *S. cerevisiae* the CUG codons are more uniformly distributed. Furthermore, the CUG distribution pattern displayed by *C. albicans* is shared among the other CTG clade species, indicating that the CUG reassignment was not a random event. It has been optimized to reduce deleterious effects of ambiguity on protein folding and structure. Interestingly, the CUG codons conserved in orthologs protein sets are mainly located in genes involved in virulence, pathogenesis or transduction, suggesting an important role for CUG ambiguity in *C. albicans* virulence. In other words, *C. albicans* CTG ambiguity forced relocation of the CUG codons to produce protein structural robustness and functional plasticity (Rocha *et al.*, 2011).

5.3. *C. albicans* and *S. cerevisiae* respond to genetic code ambiguity in different ways

The ser-tRNA$_{CAG}$ appeared at 272 ± 25 million years (My) ago in the ancestor of *C. albicans* and *S. cerevisiae* (Massey *et al.*, 2003). Approximately 171 ± 27 Million years (My) ago the *Candida* and *Saccharomyces* genera diverged, suggesting that the CUG codons at the *Saccharomyces* ancestor was ambiguous for approximately 100 My. The ser-tRNA$_{CAG}$ was maintained in *Candida* and was lost in *Saccharomyces* (Massey *et al.*, 2003;Miranda *et al.*, 2006). To clarify whether that ambiguity altered *C. albicans* biology we have compared the transcriptomes of mistranslating *C. albicans* and *S. cerevisiae*. *S. cerevisiae* mistranslating strains showed higher percentage of shared deregulated genes between them than the *C. albicans* mistranslating strains. These results are in agreement with previous studies showing higher number of deregulated and overlapped genes in *S. cerevisiae* exposed to stress than in *C. albicans* (Enjalbert *et al.*, 2006).

Comparison of the stress responses of *S. cerevisiae*, *C. albicans* and the fission yeast *Schizosaccharomyces pombe* using DNA microarrays showed that these species respond in a rather specific way to stress (Gasch, 2007). However, the highly divergent *S. cerevisiae* and *S. pombe*, respond to stress by inducing a common gene expression response, designated as the environmental stress response (ESR) (Enjalbert *et al.*, 2003;Gasch, 2007). Since *S. pombe* diverged from *Candida* and *S. cerevisiae* genera approximately 500 My ago (Berbee and Taylor, 1993) (before the appearance of the tRNA$_{CAG}^{Ser}$) and the ESR exists both in *S. pombe* and *S. cerevisiae*, the lack of the ESR in *C. albicans* suggests that CUG ambiguity may have reprogramed the stress response of the CTG clade species. It would be interesting to compare the transcriptomes of more CTG clade species mistranslating CUGs and other codons at higher level to confirm the impact of the genetic code alteration on the stress response of the CTG clade.

Furthermore, the clustering of the transcriptome profiles of the *S. cerevisiae* and *C. albicans* mistranslating strains showed that strains grouped by organism and growth conditions and not by the amino acid that was target by Ser misincorporation. The cellular processes that were most negatively affected by Ser

misincorporation were ribosome biogenesis and RNA processing. In the case of *C. albicans*, the highest level of deregulation was observed at 30ºC in strains misincorporating Ser in Leu and Ala codons. *S. cerevisiae* strains misincorporating Ser in Leu, Thr and Tyr codons showed identical trends concerning the down regulation of ribosome biogenesis processes (Paredes *et al.*, 2012). On the other hand, oxidation-reduction was up-regulated in *S. cerevisiae* strains misincorporating Ser in Leu, Thr and Tyr codons and in *C. albicans* grown at 30ºC misincorporating Ser at Ala, Leu and Lys codons.

The enrichment analysis of transcription factor motifs present in the deregulated genes showed that *C. albicans* and *S. cerevisiae* use different transcription factors to control the response to codon ambiguity. *S. cerevisiae* uses Gis1, which is involved in the response to glucose, ammonium, and phosphate starvation (Zhang *et al.*, 2009), and Pho4 which is involved in the response to inorganic phosphate availability (Zhou and O'Shea, 2011). Conversely, *C. albicans* uses the Gat1, Gln3 transcription factors which regulate nitrogen starvation-induced filamentous growth responses (Dabas and Morschhauser, 2007), Gzf3 which is involved in the oxidative stress response through the main NADPH-producing pathway, namely the pentose phosphate pathway (Wang *et al.*, 2006) and the Hap2, Hap3, Hap4 and Hap5, transcription factor motifs which are involved in the iron homeostasis (Baek *et al.*, 2008;Singh *et al.*, 2011). Future studies should focus on the transcriptional network of CUG ambiguity responses using Chip-Chip location profiling to confirm these differences between *C. albicans* and *S. cerevisiae.*

5.4. Conclusion

Our work show that the genetic code is flexible even in a highly complex eukaryotic cell. We have also shown that, within certain limits, the deleterious effects of genetic code ambiguity are dependent on growth conditions. The fluctuating levels of Leu insertion in CUG codons show that *C. albicans* is highly adapted to CUG ambiguity and that such ambiguity could be an important component of its adaptation to environmental changes. Moreover, *C. albicans* and

S. cerevisiae respond differently to genetic code ambiguity with the former failing to activate the ESR. These differences resulted from activation of different transcription factors in both species. The data shows therefore that CUG reassignment had a major impact on the CTG clade stress response.

5.5. Future Studies

Our data unveiled new questions that require future studies to be clarified, namely:

1. Quantify CUG ambiguity levels *in vivo* in mice, during infection.
2. The increase in Leu incorporation at CUG codons under stress conditions, suggest that *C. albicans* responds to stress by increasing the pool of proteins (statistical proteins), whether or not they are relevant to stress adaptation needs to be clarified in future studies, a possible approach would be using a non-ambiguous *C. albicans* strain and study is capacity to adapt to new stress.
3. Evolve ambiguous strains until tolerance to Ser misincorporation is achieved and determine whether the resulting strains reprogram the stress response during the course of such evolution.
4. It would be interesting to study the transcriptional network of Ser misincorporating strains using Chip-Chip location profiling to confirm the differences between *C. albicans* and *S. cerevisiae.*

5.6. References

Ahmad,A., Kabir,M.A., Kravets,A., Andaluz,E., Larriba,G., and Rustchenko,E. (2008). Chromosome instability and unusual features of some widely used strains of *Candida albicans*. Yeast *25*, 433-448.

Ambrogelly,A., Palioura,S., and Soll,D. (2007). Natural expansion of the genetic code. Nat. Chem. Biol. *3*, 29-35.

Bacher,J.M., de Crecy-Lagard,V., and Schimmel,P.R. (2005). Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. Proc. Natl. Acad. Sci. U. S. A *102*, 1697-1701.

Baek,Y.U., Li,M., and Davis,D.A. (2008). *Candida albicans* ferric reductases are differentially regulated in response to distinct forms of iron limitation by the Rim101 and CBF transcription factors. Eukaryot. Cell *7*, 1168-1179.

Barrell,B.G., Bankier,A.T., and Drouin,J. (1979). A different genetic code in human mitochondria. Nature *282*, 189-194.

Berbee,M.L. and Taylor,J.W. (1993). Dating the evolutionary radiations of the true fungi. *Canadian* Journal of Botany *71*, 1114-1127.

Bouchonville,K., Forche,A., Tang,K.E., Selmecki,A., and Berman,J. (2009). Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*. Eukaryot. Cell *8*, 1554-1566.

Crick,F.H. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367-379.

Dabas,N. and Morschhauser,J. (2007). Control of ammonium permease expression and filamentous growth by the GATA transcription factors GLN3 and GAT1 in *Candida albicans*. Eukaryot. Cell *6*, 875-888.

Di,G.M. (2005). The origin of the genetic code: theories and their relationships, a review. Biosystems *80*, 175-184.

Enjalbert,B., Nantel,A., and Whiteway,M. (2003). Stress-induced gene expression in *Candida albicans*: absence of a general stress response. Mol. Biol. Cell *14*, 1460-1467.

Enjalbert,B., Smith,D.A., Cornell,M.J., Alam,I., Nicholls,S., Brown,A.J., and Quinn,J. (2006). Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen *Candida albicans*. Mol. Biol. Cell *17*, 1018-1032.

Fernandez-Arenas,E., Cabezon,V., Bermejo,C., Arroyo,J., Nombela,C., ez-Orejas,R., and Gil,C. (2007). Integrated proteomics and genomics strategies bring new insight into *Candida albicans* response upon macrophage interaction. Mol. Cell Proteomics. *6*, 460-478.

Gasch,A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206.

Knight,R.D., Freeland,S.J., and Landweber,L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet *2*, 49-58.

Massey,S.E., Moura,G., Beltrao,P., Almeida,R., Garey,J.R., Tuite,M.F., and Santos,M.A. (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida spp.* Genome Res. *13*, 544-557.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A Genetic Code Alteration Is a Phenotype Diversity Generator in the Human Pathogen *Candida albicans*. PLoS. ONE. *2*, e996.

Miranda,I., Silva,R., and Santos,M.A. (2006). Evolution of the genetic code in yeasts. Yeast *23*, 203-213.

Ohama,T., Inagaki,Y., Bessho,Y., and Osawa,S. (2008). Evolving genetic code. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *84*, 58-74.

Paredes,J.A., Carreto,L., Simoes,J., Bezerra,A.R., Gomes,A.C., Santamaria,R., Kapushesky,M., Moura,G.R., and Santos,M.A. (2012). Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC. Biol. *10*, 55.

Pezo,V., Metzgar,D., Hendrickson,T.L., Waas,W.F., Hazebrouck,S., Doring,V., Marliere,P., Schimmel,P., and de Crecy-Lagard,V. (2004). Artificially ambiguous genetic code confers growth yield advantage. Proc. Natl. Acad. Sci. U. S. A *101*, 8593-8597.

Rocha,R., Pereira,P.J., Santos,M.A., and Macedo-Ribeiro,S. (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc. Natl. Acad. Sci. U. S. A *108*, 14091-14096.

Rustchenko,E. (2007). Chromosome instability in *Candida albicans*. FEMS Yeast Res. *7*, 2-11.

Santos,M.A., Cheesman,C., Costa,V., Moradas-Ferreira,P., and Tuite,M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Mol. Microbiol. *31*, 937-947.

Santos,M.A., Gomes,A.C., Santos,M.C., Carreto,L.C., and Moura,G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607-611.

Santos,M.A., Keith,G., and Tuite,M.F. (1993). Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. EMBO J. *12*, 607-616.

Selmecki,A., Forche,A., and Berman,J. (2010). Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot. Cell *9*, 991-1008.

Singh,R.P., Prasad,H.K., Sinha,I., Agarwal,N., and Natarajan,K. (2011). Cap2-HAP complex is a critical transcriptional regulator that has dual but contrasting roles in regulation of iron homeostasis in *Candida albicans*. J. Biol. Chem. *286*, 25154-25170.

Soll,D. and RajBhandary,U.L. (2006). The genetic code - thawing the 'frozen accident'. J. Biosci. *31*, 459-463.

Srinivasan,G., James,C.M., and Krzycki,J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Suzuki,T., Ueda,T., and Watanabe,K. (1997). The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122-1134.

Wang,Y., Cao,Y.Y., Jia,X.M., Cao,Y.B., Gao,P.H., Fu,X.P., Ying,K., Chen,W.S., and Jiang,Y.Y. (2006). Cap1p is involved in multiple pathways of oxidative stress response in *Candida albicans*. Free Radical Biology and Medicine *40*, 1201-1209.

Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y., and Osawa,S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. U. S. A *82*, 2306-2309.

Zhang,N., Wu,J., and Oliver,S.G. (2009). Gis1 is required for transcriptional reprogramming of carbon metabolism and the stress response during transition into stationary phase in yeast. Microbiology *155*, 1690-1698.

Zhou,X. and O'Shea,E.K. (2011). Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. Mol. Cell *42*, 826-836.

6. Annexes

Table 1: Overlap of upregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in alanine codons.

				Fold		
***C. albicans* Systematic Name**	***C. albicans* Common Name**	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	***C. albicans* 30ºC**	***C. albicans* 37ºC**	***S. cerevisiae***
CA3534	SSB1	YER103W	SSA4	0.00	1.61	1.64
CA2858	ACS2	YAL054C	ACS1	0.00	1.50	1.35
CA2595	ARG2	YJL071W	ARG2	1.86	0.00	1.25
CA4986	SER33	YIL074C	SER33	1.79	0.00	1.13
CA3132	IPF8904	YOR338W		26.01	0.00	2.09
CA6040	CPR6	YLR216C	CPR6	1.46	0.00	1.14
CA3727	HIS5.3f	YIL116W	HIS5	2.48	0.00	1.19
CA5536	IPF2837	YHR112C		2.05	0.00	1.35
CA1199	IPF10404	YBL049W	MOH1	6.01	0.00	2.74
CA0617	MET2	YNL277W	MET2	1.86	0.00	1.79
CA4379	MEP2	YNL142W	MEP2	2.58	-2.42	1.41
CA2977	HOM3	YER052C	HOM3	2.07	0.00	1.32
CA4184	SNZ1	YMR096W	SNZ1	2.15	0.00	1.65
CA4639	IPF1509	YLR089C	ALT1	2.05	0.00	1.35
CA0343	LYS1.5eoc	YIR034C	LYS1	1.78	0.00	1.48
CA5616	PET8	YNL003C	PET8	1.47	0.00	1.20
CA1732	ECM42	YMR062C	ECM40	1.80	0.00	1.35
CA1374	MIP1	YKL134C		1.22	0.00	1.05
CA0428	ILV2	YMR108W	ILV2	2.57	0.00	1.22
CA2559	HIS4	YCL030C	HIS4	1.75	0.00	1.45
CA3227	AAT21	YLR027C	AAT2	1.48	0.00	1.19
CA3073	IPF6108	YOR271C		1.56	1.23	0.00

Table 2: Overlap of downregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in alanine codons.

				Fold		
***C. albicans* Systematic Name**	***C. albicans* Common Name**	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	***C. albicans* 30ºC**	***C. albicans* 37ºC**	***S. cerevisiae***
CA2069	FUR1	YHR128W	FUR1	0.00	-1.15	-1.46
CA4413	MSH2	YOL090W	MSH2	-1.18	0.00	-1.18
CA0626	IPF3161	YNR054C	ESF2	-1.62	0.00	-1.26
CA2369	TBP1	YER148W	SPT15	-1.34	0.00	-1.28
CA0539	NUP57	YGR119C	NUP57	-1.21	0.00	-1.08
CA1478	IPF16426	YBL097W	BRN1	-1.27	0.00	-1.37
CA1206	HUB1	YNR032C-A	HUB1	-1.65	0.00	-1.24

Table 3: Overlap of upregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in leucine codons.

				Fold		
C. albicans **Systematic Name**	***C. albicans*** **Common Name**	***S. cerevisiae*** **Systematic Name**	***S. cerevisiae*** **Common Name**	***C. albicans*** **30ºC**	***C. albicans*** **37ºC**	***S. cerevisiae***
CA4536	COX13	YGL191W	COX13	1.41	1.46	1.25
CA3199	FBP1	YLR377C	FBP1	1.96	2.12	4.90
CA2719	SOD2	YHR008C	SOD2	0.00	1.96	1.36
CA5128	IPF5118	YOL032W		0.00	1.34	1.45
CA1329	MRF2	YGL143C	MRF1	-1.91	1.29	1.12
CA5164	MDH1	YKL085W	MDH1	0.00	1.62	1.98
CA4275	MDJ1	YFL016C	MDJ1	0.00	1.23	1.56
CA3479	ATP4	YPL078C	ATP4	0.00	1.71	1.22
CA2630	COX5A	YNL052W	COX5A	0.00	1.46	1.15
CA2415	MSD1	YPL104W	MSD1	1.33	0.00	1.28
CA5544	IPF5987	YPR127W		1.66	0.00	1.91
CA4286	IPF20011.5f	YGL037C	PNC1	2.27	0.00	2.71
CA5325	IPF2138	YBR137W		2.10	0.00	1.22
CA0917	RAD16	YBR114W	RAD16	5.80	0.00	1.44
CA2756	IPF18418	YHR029C		3.29	0.00	1.86
CA3840	IPF12745	YLR193C		1.95	0.00	1.24
CA2279	DUR31	YHL016C	DUR3	2.33	0.00	1.62
CA5561	SCO1	YBR037C	SCO1	2.12	0.00	1.30
CA0785	IPF14603	YJR100C		1.38	0.00	1.09
CA0892	PHO8.5	YDR481C	PHO8	1.91	0.00	2.00
CA3885	IPF19790	YDL019C	OSH2	1.75	0.00	1.47
CA5461	GSL23.5f	YGR032W	GSC2	1.42	0.00	4.32
CA1199	IPF10404	YBL049W	MOH1	5.79	0.00	9.63
CA0828	IPF17186	YDR533C	HSP31	1.56	0.00	2.32
CA3578	IPF11105	YBR046C	ZTA1	1.68	0.00	1.60
CA3890	TERT2	YLR318W	EST2	1.49	0.00	1.21
CA5868	IPF351	YMR090W		2.67	0.00	2.42
CA5336	IPF893	YJR008W		3.71	0.00	1.95
CA1611	SPO72	YNL242W	ATG2	2.72	0.00	1.95
CA5742	FUN31	YAL017W	PSK1	1.33	0.00	1.69
CA4048	IPF9207	YNL200C		2.86	0.00	2.43
CA0821	IPF16795	YNL274C		2.18	0.00	2.79
CA4543	PPE1	YHR075C	PPE1	1.99	0.00	1.60
CA4806	IPF1155	YLL029W		1.38	0.00	1.43
CA4913	CIS2	YLR299W	ECM38	1.39	0.00	1.80
CA3180	IPF12244	YKL034W	TUL1	1.74	0.00	1.42
CA4616	IPF5761	YHR176W	FMO1	1.80	0.00	1.27
CA3802	AUT7.exon2	YBL078C	ATG8	3.26	0.00	2.56
CA2896	IPF14981	YPL206C		1.49	0.00	1.41
CA2773	MGM1	YOR211C	MGM1	1.58	0.00	1.21
CA1231	IPF11817	YDL206W		1.97	0.00	1.48
CA5932	UBI4	YLL039C	UBI4	1.24	0.00	3.51
CA2606	IPF5466	YHR140W		2.45	0.00	2.77
CA2215	PFK26	YIL107C	PFK26	2.63	0.00	1.97
CA5824	NTH1	YBR001C	NTH2	5.84	0.00	2.75
CA3358	IPF19984	YOR125C	CAT5	1.71	0.00	1.24
CA5435	UFD4	YKL010C	UFD4	1.59	0.00	1.21
CA5800	IPF4175	YBR026C	ETR1	1.38	0.00	1.32
CA4709	IPF5895	YOR245C	DGA1	2.02	0.00	1.24
CA4361	IPF16104	YJR119C		2.07	0.00	1.45
CA4161	IPF20164	YJL060W	BNA3	1.48	0.00	1.31
CA4448	PHO85	YPL031C	PHO85	1.33	0.00	1.30
CA3426	IPF11869	YNL092W		1.88	0.00	1.85
CA4084	TPS1	YBR126C	TPS1	5.33	0.00	2.19
CA2561	R2	YLR438W	CAR2	1.36	0.00	3.39
CA0697	VTC4	YJL012C	VTC4	2.49	0.00	3.65
CA3566	GUT2	YIL155C	GUT2	1.32	0.00	2.94
CA1673	PST2	YDR032C	PST2	2.24	0.00	1.58
CA4255	FUM11	YPL262W	FUM1	1.25	1.18	0.00
CA2889	IPF13628	YML095C	RAD10	1.43	1.24	0.00
CA2998	LPD1	YFL018C	LPD1	1.27	1.38	0.00
CA3587	AAH1	YNL141W	AAH1	5.75	2.01	-3.21

Table 4: Overlap of downregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in leucine codons.

				Fold		
C. albicans **Systematic Name**	***C. albicans*** **Common Name**	***S. cerevisiae*** **Systematic Name**	***S. cerevisiae*** **Common Name**	***C. albicans*** **30ºC**	***C. albicans*** **37ºC**	***S. cerevisiae***
CA0733	NIP1	YMR309C	NIP1	-2.13	-2.28	-1.64
CA1875	RPB9	YGL070C	RPB9	-1.56	-1.26	-1.26
CA5822	YRB1	YDR002W	YRB1	-1.52	-1.44	-1.59
CA2471	TOP2	YNL088W	TOP2	-1.31	-1.32	-1.31
CA5644	IPF447	YNL110C	NOP15	-6.16	-1.72	-1.96
CA4010	MAK10.3	YEL053C	MAK10	-1.52	-1.44	-1.41
CA1127	MSH3	YCR092C	MSH3	0.00	-1.67	-1.24
CA0332	FEN2	YCR028C	FEN2	0.00	-1.50	-1.45
CA4296	IPF2605	YOL098C		0.00	-1.84	-1.23
CA1124	UBP12	YJL197W	UBP12	0.00	-1.34	-1.20
CA0649	RFC3	YNL290W	RFC3	0.00	-1.27	-1.32
CA5273	NUP49	YGL172W	NUP49	0.00	-1.98	-1.27
CA0887	IPF12383	YIL003W	CFD1	0.00	-1.25	-1.35
CA4496	IPF4062	YDR333C		1.25	-1.12	-1.20
CA0886	IPF12382	YER002W	NOP16	-5.75	0.00	-2.12
CA1275	RPC53	YDL150W	RPC53	-3.06	0.00	-1.62
CA5719	IPF2441	YER126C	NSA2	-6.04	0.00	-2.08
CA1438	NOP58	YOR310C	NOP58	-2.96	0.00	-1.93
CA2573	SUI2	YJR007W	SUI2	-1.90	0.00	-1.61
CA0730	IPF7334	YOR004W		-2.30	0.00	-1.79
CA6076	PAC10.3	YGR078C	PAC10	-1.52	0.00	-1.52
CA2911	MET7	YOR241W	MET7	-1.32	0.00	-1.19
CA5902	SRP68	YPL243W	SRP68	-2.00	0.00	-1.37
CA3782	SER2	YGR208W	SER2	-1.77	1.43	-1.32
CA1063	NMT1	YLR195C	NMT1	-1.70	0.00	-1.27
CA3491	RPC40	YPR110C	RPC40	-3.23	0.00	-1.82
CA0088	IPF7349	YNL256W	FOL1	-2.77	0.00	-1.77
CA6021	BFR2	YDR299W	BFR2	-4.21	0.00	-1.81
CA3500	IPF14487	YDR361C	BCP1	-2.43	0.00	-1.98
CA5359	TIF35	YDR429C	TIF35	-2.02	0.00	-1.70
CA2757	IPF19767	YPR097W		-2.06	0.00	-1.36
CA4506	GCD10	YNL062C	GCD10	-2.97	0.00	-1.73
CA3237	SPB4	YFL002C	SPB4	-4.85	0.00	-1.39
CA3663	IPF11432	YJR072C	NPA3	-2.04	0.00	-1.48
CA4811	ECM16	YMR128W	ECM16	-6.57	0.00	-2.03
CA3562	IPF7823	YDR190C	RVB1	-1.70	0.00	-1.35
CA2682	FRS1	YLR060W	FRS1	-1.38	0.00	-1.45
CA1736	IPF4983	YNL035C		-3.13	0.00	-1.50
CA1236	YPT6	YLR262C	YPT6	-2.11	0.00	-1.40
CA2252	IPF12844	YOR287C		-3.87	0.00	-2.04
CA3892	PSE1	YMR308C	PSE1	-2.20	0.00	-1.71
CA1041	BMS1	YPL217C	BMS1	-3.84	0.00	-2.19
CA2304	PRP43	YGL120C	PRP43	-5.56	0.00	-1.86
CA2480	CCT7	YJL111W	CCT7	-1.94	0.00	-1.46
CA3592	IPF12900	YPL199C		-1.79	0.00	-1.31
CA1598	SES1	YDR023W	SES1	-1.89	0.00	-1.45
CA0949	IPF10269	YNL153C	GIM3	-1.42	0.00	-1.67
CA5935	IPF2798	YNL175C	NOP13	-5.61	0.00	-2.25
CA5882	RRP1	YDR087C	RRP1	-7.46	0.00	-2.08
CA6072	DBP10	YDL031W	DBP10	-5.54	0.00	-1.71
CA3893	FUN11	YAL036C	RBG1	-2.16	0.00	-1.78
CA3075	IPF6105	YMR217W	GUA1	-2.40	0.00	-1.81
CA1443	IPF4776	YKL172W	EBP2	-6.86	0.00	-2.11
CA5507	TIF11	YMR260C	TIF11	-1.68	0.00	-1.61
CA5957	RRP3.3eoc	YHR065C	RRP3	-5.04	0.00	-1.44
CA3348	RIO1	YOR119C	RIO1	-4.39	0.00	-1.61
CA4326	IPF2093	YGR103W	NOP7	-2.60	0.00	-2.38
CA3508	RSA2	YMR131C	RRB1	-3.48	0.00	-1.99
CA4211	IPF5052	YOL041C	NOP12	-4.56	0.00	-1.79
CA2400	SUP45	YBR143C	SUP45	-1.98	0.00	-1.42
CA5833	IMP3	YHR148W	IMP3	-4.86	0.00	-1.95
CA4096	KRR1	YCL059C	KRR1	-6.76	0.00	-2.40

				Fold		
***C. albicans* Systematic Name**	***C. albicans* Common Name**	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	***C. albicans* 30ºC**	***C. albicans* 37ºC**	***S. cerevisiae***
CA1306	CDC14.3	YFR028C	CDC14	-1.55	0.00	-1.34
CA3242	RAT1	YOR048C	RAT1	-1.81	0.00	-1.59
CA3010	IPF14568	YGL246C	RAI1	-3.01	0.00	-1.36
CA1161	GRX3	YDR098C	GRX3	-1.54	0.00	-1.39
CA5568	ELF1	YPL226W	NEW1	-4.75	0.00	-1.73
CA2592	IPF18447	YBR267W	REI1	-6.93	0.00	-1.64
CA2966	RPC34	YNR003C	RPC34	-3.64	0.00	-1.67
CA5963	IPF966	YML093W	UTP14	-5.49	0.00	-2.09
CA2888	GIM5	YML094W	GIM5	-1.37	0.00	-1.45
CA1417	IPF16126	YNL113W	RPC19	-6.90	0.00	-1.77
CA5869	MIS12	YBR084W	MIS1	-2.14	0.00	-1.81
CA2012	SEC10	YLR166C	SEC10	-1.08	0.00	-1.10
CA4247	TAF90	YBR198C	TAF5	-1.28	0.00	-1.19
CA0048	TIF4631	YGR162W	TIF4631	-1.67	0.00	-1.92
CA5993	ASP1	YDR321W	ASP1	-3.70	0.00	-1.83
CA3584	IPF15468	YCR051W		-1.92	0.00	-1.66
CA0294	IPF16061	YOR243C	PUS7	-3.21	0.00	-1.93
CA0098	IPF16479	YDR365C	ESF1	-4.74	0.00	-2.36
CA2399	MAK5	YBR142W	MAK5	-4.03	0.00	-2.22
CA1941	IPF7998	YLR009W	RLP24	-6.32	0.00	-1.90
CA0466	URA4	YLR420W	URA4	-1.63	0.00	-2.11
CA2662	IPF17139	YKR081C	RPF2	-5.06	0.00	-2.36
CA3894	YIF2	YAL035W	FUN12	-1.77	0.00	-1.45
CA1480	HIS6	YIL020C	HIS6	-1.34	0.00	-1.41
CA1630	RPP1	YHR062C	RPP1	-2.57	0.00	-1.75
CA0245	POL5	YEL055C	POL5	-1.72	0.00	-1.62
CA1770	IPF12457	YGL111W	NSA1	-6.20	0.00	-2.00
CA0866	IPF14775	YNR046W		-3.07	0.00	-1.57
CA2002	IPF16470	YBL036C		-2.90	0.00	-1.54
CA0667	TIF5	YPR041W	TIF5	-2.77	0.00	-1.47
CA2762	IPF15012	YBL032W	HEK2	-2.52	0.00	-1.45
CA2814	IPF5607	YLR435W	TSR2	-4.36	0.00	-1.88
CA6098	IPF61	YDL148C	NOP14	-8.88	0.00	-2.23
CA0443	SPB1	YCL054W	SPB1	-4.90	0.00	-1.99
CA2261	RAS1	YOR101W	RAS1	-1.57	0.00	-1.65
CA2373	CBF5	YLR175W	CBF5	-1.92	0.00	-1.75
CA1274	SAS10	YDL153C	SAS10	-6.35	0.00	-2.27
CA5297	YAR1	YPL239W	YAR1	-2.20	0.00	-1.56
CA5149	HGH1	YGR187C	HGH1	-2.51	0.00	-1.53
CA1427	IPF9914	YOR335C	ALA1	-1.71	0.00	-1.43
CA4089	IPF5009	YGR245C	SDA1	-4.59	0.00	-2.39
CA4892	PRT1	YOR361C	PRT1	-2.48	0.00	-1.58
CA5323	IPF2142	YLR002C	NOC3	-6.33	0.00	-1.52
CA5992	FAA4	YMR246W	FAA4	-1.83	0.00	-1.48
CA2600	IPF11177	YMR012W	CLU1	-2.26	0.00	-1.26
CA1500	IPF13221	YOR091W		-3.22	0.00	-1.68
CA3499	RPG1	YBR079C	RPG1	-1.96	0.00	-1.81
CA3631	TIM54	YJL054W	TIM54	-2.03	0.00	-1.38
CA4553	CYB1	YPR119W	CLB2	-1.40	0.00	-1.74
CA4003	IPF4257	YDR324C	UTP4	-7.22	0.00	-1.83
CA3007	IPF12148	YNL182C	IPI3	-6.33	0.00	-1.88
CA1666	SRP40	YKR092C	SRP40	-7.36	0.00	-1.95
CA3784	IPF11229	YCL031C	RRP7	-2.16	0.00	-1.52
CA5682	NOG1	YPL093W	NOG1	-4.10	0.00	-2.00
CA3494	RPB5	YBR154C	RPB5	-2.12	0.00	-1.83
CA1747	IPF10884	YLR003C		-7.21	0.00	-1.78
CA3401	IPF8652	YGR173W	RBG2	-1.86	0.00	-1.74
CA0813	MPP10	YJR002W	MPP10	-4.06	0.00	-2.01

				Fold		
C. albicans **Systematic Name**	***C. albicans*** **Common Name**	***S. cerevisiae*** **Systematic Name**	***S. cerevisiae*** **Common Name**	***C. albicans*** **30ºC**	***C. albicans*** **37ºC**	***S. cerevisiae***
CA1837	IPF5776	YLR015W	BRE2	-1.74	0.00	-1.51
CA4718	IPF8661	YOR056C	NOB1	-7.21	0.00	-1.79
CA0470	YPT32	YER031C	YPT31	-1.27	0.00	-1.35
CA3830	SSF1		SSF1	-5.88	0.00	-1.70
CA5531	GCD14	YJL125C	GCD14	-3.62	0.00	-1.74
CA3868	CCT1	YDR212W	TCP1	-1.80	0.00	-1.50
CA3477	IPF3229	YIL064W		-2.88	0.00	-2.27
CA4732	IPF7733	YCR063W	BUD31	-1.82	0.00	-1.42
CA3006	IPF12152	YDL213C	NOP6	-4.54	0.00	-2.34
CA2629	IPF10911	YIL110W		-7.29	0.00	-1.63
CA1106	IPF11270	YOL077C	BRX1	-5.08	0.00	-1.95
CA1803	IPF16748	YER006W	NUG1	-4.89	0.00	-2.13
CA2449	ENP1	YBR247C	ENP1	-11.13	0.00	-1.98
CA4928	IPF1399	YDR346C	SVF1	-1.76	0.00	-1.65
CA0599	IPF12428	YHR088W	RPF1	-5.29	0.00	-1.69
CA2283	MRT4	YKL009W	MRT4	-4.46	0.00	-2.05
CA2992	RRS1	YOR294W	RRS1	-5.68	0.00	-1.96
CA3497	FEN11	YCR034W	FEN1	-2.61	0.00	-1.56
CA6134	MAK21	YDR060W	MAK21	-3.08	0.00	-1.91
CA5457	TRM1	YDR120C	TRM1	-5.64	0.00	-1.95
CA6016	IPF630	YOR145C	PNO1	-5.32	0.00	-1.64
CA1277	IPF10837	YDR083W	RRP8	-3.01	0.00	-2.42
CA1784	ARO7	YPR060C	ARO7	-1.99	0.00	-1.58
CA0119	RMT2	YDR465C	RMT2	-2.45	0.00	-2.08
CA0052	IPF19567	YHR085W	IPI1	-3.69	0.00	-1.60
CA0923	IPF12987	YMR049C	ERB1	-3.40	0.00	-1.93
CA1293	H4	YJL033W	HCA4	-6.20	0.00	-2.12
CA0020	IPF7046	YHR052W	CIC1	-10.68	0.00	-1.74
CA4065	RPC25	YKL144C	RPC25	-2.62	0.00	-1.48
CA4289	IPF6665	YGR145W	ENP2	-8.77	0.00	-2.28
CA2414	IPF13756.3f	YJL010C		-5.49	0.00	-2.01
CA2393	RHO3	YIL118W	RHO3	-1.42	0.00	-1.20
CA5292	IPF1948	YJL069C	UTP18	-7.22	0.00	-1.48
CA3927	DIP2	YLR129W	DIP2	-4.78	0.00	-1.68
CA1055	IPF17504	YBL055C		-1.40	0.00	-1.59
CA1479	IPF16428	YIL019W	FAF1	-2.27	0.00	-2.62
CA2311	SSO2	YMR183C	SSO2	-1.29	0.00	-1.40
CA2881	RAD4	YOR206W	NOC2	-4.45	0.00	-1.95
CA4156	IPF10990	YOR006C		-4.67	0.00	-1.63
CA2099	IPF11484	YKL099C	UTP11	-3.08	0.00	-2.06
CA4367	MNN10	YDR245W	MNN10	-1.49	0.00	-1.45
CA2461	MAK11	YKL021C	MAK11	-5.50	0.00	-2.02
CA1202	DBP3	YGL078C	DBP3	-5.34	0.00	-1.92
CA1206	HUB1	YNR032C-A	HUB1	-1.55	0.00	-1.26
CA0298	PPT1	YGR123C	PPT1	-5.52	0.00	-1.90
CA4575	IPF9428	YPL146C		-7.20	0.00	-2.36
CA1002	ROK1.3	YGL171W	ROK1	-3.47	0.00	-2.25
CA0912	RPA49	YNL248C	RPA49	-3.39	0.00	-2.04
CA5929	GCD2	YGR083C	GCD2	-1.86	0.00	-1.52
CA2663	NUP133	YKR082W	NUP133	-1.61	0.00	-1.23
CA1291	IPF16752	YGR200C	ELP2	-2.72	0.00	-1.73
CA2030	IPF13116	YGL099W	LSG1	-5.17	0.00	-1.84
CA5563	PHO88	YBR106W	PHO88	-1.96	0.00	-1.32
CA2003	IPF16471	YDL051W	LHP1	-2.27	0.00	-2.24
CA2035	SGD1.5f	YLR336C	SGD1	-3.57	0.00	-1.66
CA0852	KIP2	YPL155C	KIP2	-2.81	-2.04	0.00
CA5346	IPF867	YKR029C	SET3	-1.22	-1.99	0.00
CA5217	RFC2	YJR068W	RFC2	-1.50	-1.34	0.00

				Fold		
C. albicans Systematic Name	*C. albicans* Common Name	*S. cerevisiae* Systematic Name	*S. cerevisiae* Common Name	*C. albicans* 30ºC	*C. albicans* 37ºC	*S. cerevisiae*
CA0205	DBF2	YGR092W	DBF2	-1.82	-1.74	0.00
CA0280	IPF10566	YDR315C	IPK1	-2.11	-1.62	0.00
CA5305	IPF2175	YMR161W	HLJ1	-1.27	-1.16	0.00
CA2901	PUS1	YPL212C	PUS1	-3.40	-1.20	0.00
CA0846	SMC3	YJL074C	SMC3	-3.01	-2.29	0.00
CA1095	SMC1	YFL008W	SMC1	-3.39	-1.71	0.00
CA3754	TIF3	YPR163C	TIF3	-2.33	-1.75	0.00
CA4554	IPF7874	YHR154W	RTT107	-2.20	-2.46	0.00
CA0723	PRI2	YKL045W	PRI2	-6.76	-2.61	0.00

Table 5: Overlap of upregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in lysine codons.

				Fold		
C. albicans Systematic Name	*C. albicans* Common Name	*S. cerevisiae* Systematic Name	*S. cerevisiae* Common Name	*C. albicans* 30ºC	*C. albicans* 37ºC	*S. cerevisiae*
CA5547	IPF5981	YML128C	MSC1	0.00	2.31	3.17
CA4974	IPF4704	YJR111C		0.00	1.49	1.23
CA1492	IPF168	YDL046W	NPC2	0.00	1.48	1.49
CA2016	YIM1	YMR152W	YIM1	0.00	1.25	1.17
CA1673	PST2	YDR032C	PST2	0.00	2.06	1.45

Table 6: Overlap of downregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in lysine codons.

				Fold		
C. albicans Systematic Name	*C. albicans* Common Name	*S. cerevisiae* Systematic Name	*S. cerevisiae* Common Name	*C. albicans* 30ºC	*C. albicans* 37ºC	*S. cerevisiae*
CA2069	FUR1	YHR128W	FUR1	0.00	-1.34	-1.10
CA3727	HIS5.3f	YIL116W	HIS5	0.00	-2.05	-1.07
CA4734	TFA1	YKL028W	TFA1	0.00	-1.11	-1.72

Table7: Overlap of upregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in threonine codons.

				Fold		
C. albicans Systematic Name	*C. albicans* Common Name	*S. cerevisiae* Systematic Name	*S. cerevisiae* Common Name	*C. albicans* 30ºC	*C. albicans* 37ºC	*S. cerevisiae*
CA5263	CYB2	YML054C	CYB2	0.00	2.26	2.21
CA4159	ALD5	YOR374W	ALD4	0.00	1.99	3.09
CA0828	IPF17186	YDR533C	HSP31	0.00	2.01	1.78
CA3132	IPF8904	YOR338W		15.14	0.00	1.89
CA2896	IPF14981	PGC1		1.42	0.00	1.61

Table 8: Overlap of downregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in threonine codons.

				Fold		
C. albicans Systematic Name	*C. albicans* Common Name	*S. cerevisiae* Systematic Name	*S. cerevisiae* Common Name	*C. albicans* 30ºC	*C. albicans* 37ºC	*S. cerevisiae*
CA6031	PRO1	YDR300C	PRO1	0.00	-1.90	-1.47
CA3579	HSH49	YOR319W	HSH49	0.00	-2.29	-1.46
CA2682	FRS1	YLR060W	FRS1	-1.32	0.00	-1.42
CA1475	GLE2	YER107C	GLE2	-1.25	0.00	-1.38

Table 9: Overlap of upregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in tyrosine codons.

				Fold		
***C. albicans* Systematic Name**	***C. albicans* Common Name**	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	***C. albicans* 30ºC**	***C. albicans* 37ºC**	***S. cerevisiae***
CA3381	NPR2	YEL062W	NPR2	0.00	1.08	1.13
CA3727	HIS5.3f	YIL116W	HIS5	0.00	1.58	1.32
CA6166	CTP1	YBR291C	CTP1	0.00	1.35	1.23
CA3534	SSB1	YER103W	SSA4	0.00	1.53	3.66
CA2294	IPF19946	YGL091C	NBP35	0.00	1.14	1.18
CA5868	IPF351	YMR090W		0.00	1.55	2.03
CA1932	IPF3903	YPL183W-A		0.00	1.21	1.32
CA5164	MDH1	YKL085W	MDH1	0.00	1.28	1.59
CA1484	ARO4	YBR249C	ARO4	0.00	1.32	1.23
CA0428	ILV2	YMR108W	ILV2	0.00	1.40	1.57
CA1015	IPF15547	YMR293C		0.00	1.33	1.18
CA2559	HIS4	YCL030C	HIS4	0.00	1.42	1.45
CA4286	IPF20011.5f	YGL037C	PNC1	1.96	0.00	2.12
CA3885	IPF19790	YDL019C	OSH2	1.80	0.00	1.42
CA5309	IPF2170	YEL030W	ECM10	1.17	0.00	1.49
CA0821	IPF16795	YNL274C		1.80	0.00	2.20
CA4806	IPF1155	YLL029W		1.35	0.00	1.41
CA2773	MGM1	YOR211C	MGM1	1.70	0.00	1.26
CA2777	THI6	YPL214C	THI6	1.38	0.00	1.28
CA3524	LKH1.3	YNL045W		1.20	0.00	1.29
CA4361	IPF16104	YJR119C		1.79	0.00	1.46
CA2872	APG13	YPR185W	ATG13	1.70	1.37	0.00

Table 10: Overlap of downregulated orthologous genes in *S. cerevisiae* and *C.albicans* strains grown at 30ºC and 37ºC, misincorporating serine in tyrosine codons.

				Fold		
***C. albicans* Systematic Name**	***C. albicans* Common Name**	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	***C. albicans* 30ºC**	***C. albicans* 37ºC**	***S. cerevisiae***
CA3562	IPF7823	YDR190C	RVB1	0.00	-1.22	-1.23
CA4296	IPF2605	YOL098C		0.00	-1.57	-1.10
CA2471	TOP2	YNL088W	TOP2	0.00	-1.28	-1.20
CA5281	SPT6	YGR116W	SPT6	0.00	-1.14	-1.18
CA5560	THI13	YDL244W	THI13	0.00	-1.27	-2.01
CA5549	ERG13	YML126C	ERG13	-1.77	0.00	-1.22
CA0949	IPF10269	YNL153C	GIM3	-1.24	0.00	-1.29
CA5507	TIF11	YMR260C	TIF11	-1.66	0.00	-1.36
CA3401	IPF8652	YGR173W	RBG2	-1.82	0.00	-1.45
CA5364	DCD1	YHR144C	DCD1	-1.20	0.00	-1.19
CA4760	IPF9652	YJL011C	RPC17	-2.43	0.00	-1.15
CA0450	IPF9353	YMR123W	PKR1	-2.23	0.00	-1.27
CA3853	MVD1.3	YNR043W	MVD1	-1.42	0.00	-1.30

Table 11: GO enrichment for the different groups of dowregulated genes identified in the transcriptome profile of *C. albicans* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
27	ribosomal large subunit assembly	30			14	1.30E-04						
278	mitotic cell cycle	188			39	1.51E-03						
460	maturation of 5.8S rRNA	61			31	1.25E-13						
469	cleavage involved in rRNA processing	54			23	1.18E-07						
478	endonucleolytic cleavage involved in rRNA processing	35			18	2.78E-07						
967	rRNA 5'-end processing	27			16	1.63E-07						
6139	nucleobase-containing compound metabolic process	1065			170	5.56E-11						
6261	DNA-dependent DNA replication	75			19	4.71E-02						
6364	rRNA processing	173			69	1.10E-24						
6396	RNA processing	375			95	1.09E-17						
6412	translation	412									14	4.29E-03
6807	nitrogen compound metabolic process	1314			199	3.79E-11						
7062	sister chromatid cohesion	35			15	1.90E-04						
7064	mitotic sister chromatid cohesion	27			13	2.40E-04						
9059	macromolecule biosynthetic process	843									20	1.13E-02
9987	cellular process	3766			421	6.24E-06						
10467	gene expression	971			153	5.15E-09					24	6.70E-04
16070	RNA metabolic process	638			125	6.08E-14						
16072	rRNA metabolic process	180			69	1.89E-23						
22613	ribonucleoprotein complex biogenesis	296	12	1.68E-02	91	2.18E-23						
30490	maturation of SSU-rRNA	77			30	2.68E-09						
32259	methylation	57			16	4.88E-02						
34470	ncRNA processing	260			85	9.59E-24						
34660	ncRNA metabolic process	308			88	5.06E-20						
42254	ribosome biogenesis	259	11	2.28E-02	90	1.29E-27						
42493	response to drug	349			56	4.00E-02						
42797	tRNA transcription from RNA polymerase III promoter	16			8	4.03E-02						
43170	macromolecule metabolic process	1915			240	4.25E-05						
44237	cellular metabolic process	2641			303	2.32E-03						
44238	primary metabolic process	2525			296	3.80E-04						
44249	cellular biosynthetic process	1268									24	6.10E-02
70925	organelle assembly	73			20	8.39E-03						
71840	cellular component organization or biogenesis	1360			192	1.34E-07						
90304	nucleic acid metabolic process	882			159	4.83E-15						
90305	nucleic acid phosphodiester bond hydrolysis	94			26	3.00E-04						

Table 12: GO enrichment for the different groups of upregulated genes identified in the transcriptome profile of *C. albicans* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
96	sulfur amino acid metabolic process	34	8	7.21E-03								
6082	organic acid metabolic process	321	45	1.49E-13								
6520	cellular amino acid metabolic process	225	41	2.36E-16								
6525	arginine metabolic process	13	7	3.68E-05								
6551	leucine metabolic process	7	5	4.20E-04								
6553	lysine metabolic process	8	4	3.90E-02								
6555	methionine metabolic process	23	7	3.88E-03								
6790	sulfur compound metabolic process	64	10	2.89E-02								
9064	glutamine family amino acid metabolic process	34	9	7.20E-04								
9066	aspartate family amino acid metabolic process	40	14	1.33E-08								
9081	branched chain family amino acid metabolic process	16	10	6.79E-09								
9099	valine biosynthetic process	3	3	1.79E-02								
9308	amine metabolic process	261	41	6.28E-14								
15837	amine transport	59	10	1.40E-02								
15849	organic acid transport	78	15	2.60E-05								
19740	nitrogen utilization	12	5	1.40E-02								
42180	cellular ketone metabolic process	332	46	1.05E-13								
43436	oxoacid metabolic process	230	45	1.32E-13								
44271	cellular nitrogen compound biosynthetic process	270	38	3.66E-11								
44281	small molecule metabolic process	665	54	8.74E-07								
55085	transmembrane transport	329	39	4.40E-09								
55114	oxidation-reduction process	450	33	1.66E-02	74	4.30E-04	3	9.94E-03				
71702	organic substance transport	153	21	2.97E-05								

Table 13: GO enrichment for the different groups of downregulated genes identified in the transcriptome profile of *C. albicans* strains grown at 37ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
162	tryptophan biosynthetic process	6					3	1.22E-02				
35672	oligopeptide transmembrane transport	6							2	8.16E-02		
6566	threonine metabolic process	7					4	3.00E-04				
9082	branched chain family amino acid biosynthetic process	12					4	4.11E-03				
19740	nitrogen utilization	12	2	2.06E-02								
9073	aromatic amino acid family biosynthetic process	13					4	5.87E-03				
9086	methionine biosynthetic process	20					4	3.68E-02				
19438	aromatic compound biosynthetic process	29					5	1.16E-02				
9066	aspartate family amino acid metabolic process	40					10	3.86E-08				
8652	cellular amino acid biosynthetic process	124					20	1.30E-13				
9309	amine biosynthetic process	131					20	3.95E-13				
16053	organic acid biosynthetic process	163					20	2.95E-11				
6520	cellular amino acid metabolic process	225					20	1.28E-08				
44271	cellular nitrogen compound biosynthetic process	270					22	6.86E-09				
55085	transmembrane transport	329	5	4.03E-02					8	9.55E-02		
44281	small molecule metabolic process	665					24	5.44E-03				
55146	oxidation-reduction process	450							8	9.55E-02		

Table 14: GO enrichment for the different groups of upregulated genes identified in the transcriptome profile of *C. albicans* strains grown at 37ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
27	ribosomal large subunit assembly	30	6	4.60E-04								
41	transition metal ion transport	36			4	4.16E-02						
955	amino acid catabolic process via Ehrlich pathway	6									3	5.22E-03
5996	monosaccharide metabolic process	72	8	8.30E-04								
6006	glucose metabolic process	38	7	1.00E-04								
6007	glucose catabolic process	29	7	1.44E-05							4	5.00E-02
6086	acetyl-CoA biosynthetic process from pyruvate	4			2	7.37E-02						
6091	generation of precursor metabolites and energy	130	8	6.29E-02	13	8.35E-09					11	1.31E-05
6094	gluconeogenesis	8	3	3.09E-02								
6096	glycolysis	17	6	1.10E-05								
6119	oxidative phosphorylation	31			8	2.30E-08					7	4.33E-06
6120	mitochondrial electron transport, NADH to ubiquinone	7			3	3.58E-03					5	4.92E-07
6123	mitochondrial electron transport, cytochrome c to oxygen	11			3	1.65E-02						
6412	translation	412	31	1.28E-14								
6448	regulation of translational elongation	18	6	1.63E-05								
6450	regulation of translational fidelity	12	4	3.21E-03								

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
6520	cellular amino acid metabolic process	225									10	1.88E-02
6558	L-phenylalanine metabolic process	3									2	8.22E-02
6559	L-phenylalanine catabolic process	2									2	2.76E-02
6568	tryptophan metabolic process	10									3	3.04E-02
6586	indolalkylamine metabolic process	10									3	3.04E-02
6811	ion transport	171			8	1.74E-02						
6812	cation transport	123			7	1.44E-02						
6825	copper ion transport	11			3	1.65E-02						
6826	iron ion transport	20			4	3.79E-03						
8152	metabolic process	2962	54	4.40E-02								
8652	cellular amino acid biosynthetic process	124									7	5.99E-02
9058	biosynthetic process	1292	39	3.62E-06								
9059	macromolecule biosynthetic process	843	33	1.36E-07								
9060	aerobic respiration	69			7	3.10E-04						
9072	aromatic amino acid family metabolic process	22									4	1.62E-02
9225	nucleotide-sugar metabolic process	8							2	8.40E-02		
9308	amine metabolic process	261									10	6.44E-02
10467	gene expression	971	32	2.11E-05								
10608	posttranscriptional regulation of gene expression	85	8	2.94E-03								
15980	energy derivation by oxidation of organic compounds	112			12	2.24E-08					9	3.80E-04
15985	energy coupled proton transport, down electrochemical gradient	17			3	6.54E-02						
15986	ATP synthesis coupled proton transport	17			3	6.54E-02						
16052	carbohydrate catabolic process	65	8	3.70E-04								
16310	phosphorylation	171			8	1.74E-02						
19319	hexose biosynthetic process	11	4	2.16E-03								
19320	hexose catabolic process	35	7	5.87E-05								
19538	protein metabolic process	1176	36	1.37E-05								
19655	glucose catabolic process to ethanol	3									2	8.22E-02
22613	ribonucleoprotein complex biogenesis	296	15	9.60E-04								
22618	ribonucleoprotein complex assembly	87	8	3.50E-03								
22900	electron transport chain	36			8	8.54E-08					7	1.32E-05
22904	respiratory electron transport chain	33			8	4.00E-08					7	6.93E-06
31347	regulation of defense response	26	6	1.80E-04								
34220	ion transmembrane transport	67			6	3.74E-03						

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
35821	modification of morphology or physiology of other organism	27	6	2.30E-04								
42254	ribosome biogenesis	259	15	1.70E-04								
42255	ribosome assembly	49	8	3.97E-05								
42273	ribosomal large subunit biogenesis	68	6	5.84E-02								
42430	indole-containing compound metabolic process	10									3	3.04E-02
42773	ATP synthesis coupled electron transport	31			8	2.30E-08					7	4.33E-06
42775	mitochondrial ATP synthesis coupled electron transport	30			8	1.72E-08					7	3.38E-06
42776	mitochondrial ATP synthesis coupled proton transport	2			2	1.24E-02						
44106	cellular amine metabolic process	241									10	3.34E-02
44237	cellular metabolic process	2641	51	1.69E-02								
44271	cellular nitrogen compound biosynthetic process	270									10	8.47E-02
44275	cellular carbohydrate catabolic process	59	8	1.70E-04								
44282	small molecule catabolic process	106	8	1.49E-02								
44283	small molecule biosynthetic process	239									11	5.61E-03
45333	cellular respiration	92			12	2.10E-09					7	8.95E-03
45727	positive regulation of translation	21	4	3.56E-02								
45901	positive regulation of translational elongation	2	2	4.70E-02								
46164	alcohol catabolic process	50	7	7.40E-04								
46364	monosaccharide biosynthetic process	20	4	2.91E-02								
46365	monosaccharide catabolic process	42	7	2.10E-04								
48584	positive regulation of response to stimulus	37	6	1.69E-03								
51701	interaction with host	90	8	4.50E-03								
55085	transmembrane transport	329			11	1.74E-02						
55114	oxidation-reduction process	450			20	7.39E-08						
65008	regulation of biological quality	340	14	2.21E-02								
70925	organelle assembly	73	8	9.20E-04								
71826	ribonucleoprotein complex subunit organization	91	8	4.88E-03								
71843	cellular component biogenesis at cellular level	388	17	1.42E-03								
75136	response to host	39	6	2.32E-03								
80134	regulation of response to stress	48	6	7.99E-03								

Table 15: GO enrichment for the different groups of downregulated genes identified in the transcriptome profile of *S. cerevisiae* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
27	ribosomal large subunit assembly	33			23	9.81E-20			7	2.62E-03	15	8.81E-11
54	ribosomal subunit export from nucleus	41			18	3.32E-10			8	1.11E-03	12	9.76E-06
460	maturation of 5.8S rRNA	81			51	9.93E-43			22	7.27E-16	33	9.54E-24
469	cleavage involved in rRNA processing	69			43	2.34E-35			19	1.32E-13	25	4.44E-16
470	maturation of LSU-rRNA	25			18	1.74E-15			7	3.40E-04	13	4.16E-10
478	endonucleolytic cleavage involved in rRNA processing	47			33	2.80E-29			15	1.72E-11	20	3.52E-14
966	RNA 5'-end processing	36			25	1.53E-21			11	8.05E-08	15	4.26E-10
2181	cytoplasmic translation	173			36	4.69E-10						
3006	developmental process involved in reproduction	160									19	4.17E-03
6139	nucleobase-containing compound metabolic process	1683			199	8.60E-33			90	3.80E-16	104	9.52E-06
6360	transcription from RNA polymerase I promoter	63			15	4.40E-04			9	3.93E-03		
6364	rRNA processing	294			127	5.13E-86			60	3.46E-40	73	1.10E-38
6396	RNA processing	567			155	2.72E-73			70	8.05E-33	78	9.94E-23
6399	tRNA metabolic process	157			35	1.15E-10			16	1.60E-04		
6400	tRNA modification	70			23	3.03E-10			9	9.43E-03		
6403	RNA localization	114			20	1.22E-03						
6413	translational initiation	58			14	8.60E-04						
6725	cellular aromatic compound metabolic process	76									14	3.60E-04
6767	water-soluble vitamin metabolic process	51	4	7.53E-03							13	1.56E-05
6807	nitrogen compound metabolic process	1951			210	4.92E-29			91	2.47E-12	122	7.28E-08
6913	nucleocytoplasmic transport	161			34	1.35E-09						
8033	tRNA processing	102			26	6.51E-09			11	6.53E-03		
8614	pyridoxine metabolic process	7									5	7.10E-04
9110	vitamin biosynthetic process	47									12	5.26E-05
9272	fungal-type cell wall biogenesis	67									12	3.06E-03
9451	RNA modification	156			31	6.89E-08						
10467	gene expression	1847			223	7.89E-41			98	2.82E-18		
10927	cellular component assembly involved in morphogenesis	62									14	2.45E-05
16070	RNA metabolic process	1237			184	5.11E-44			84	1.63E-21	87	5.16E-07
16072	rRNA metabolic process	306			129	6.86E-86			61	2.57E-40	73	2.19E-37
18130	heterocycle biosynthetic process	152									20	4.80E-04
19438	aromatic compound biosynthetic process	37									9	3.30E-03
22413	reproductive process in single-celled organism	146									20	2.50E-04
22613	ribonucleoprotein complex biogenesis	475			179	4.05E-115			80	1.55E-49	97	3.30E-45
22618	ribonucleoprotein complex assembly	129			34	1.24E-12			13	2.36E-03	20	3.07E-05
30435	sporulation resulting in formation of a cellular spore	128									17	2.77E-03
30437	ascospore formation	108									17	2.50E-04
30488	tRNA methylation	19			10	6.07E-06						
30490	maturation of SSU-rRNA	104			60	1.68E-47			28	1.40E-20	33	1.15E-19

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
31120	snRNA pseudouridine synthesis	7			5	4.89E-03						
31125	rRNA 3'-end processing	22			8	6.11E-03						
32259	methylation	109			20	5.80E-04						
32989	cellular component morphogenesis	86									15	3.00E-04
33750	ribosome localization	41			18	3.32E-10			8	1.11E-03	12	9.76E-06
34293	sexual sporulation	113									17	4.80E-04
34470	ncRNA processing	396			153	1.69E-97			70	1.34E-43	78	7.42E-34
34660	ncRNA metabolic process	451			158	1.73E-93			71	9.89E-41	78	1.11E-29
40031	snRNA modification	7			5	4.89E-03						
42254	ribosome biogenesis	405			177	2.65E-127	13	8.74E-03	78	1.06E-52	96	6.30E-51
42255	ribosome assembly	56			31	1.09E-22			10	1.50E-04	18	4.17E-10
42273	ribosomal large subunit biogenesis	81			62	3.56E-61			22	7.27E-16	40	2.13E-33
42274	ribosomal small subunit biogenesis	124			72	5.25E-58	12	2.67E-03	36	1.98E-28	39	1.90E-23
42723	thiamine-containing compound metabolic process	20									9	7.85E-06
42797	tRNA transcription from RNA polymerase III promoter	16			11	1.81E-08			5	7.44E-03		
42816	vitamin B6 metabolic process	9									5	3.98E-03
43170	macromolecule metabolic process	2827			251	1.60E-23			109	2.76E-10		
43628	ncRNA 3'-end processing	43			14	1.38E-05						
43934	sporulation	131									17	3.81E-03
43935	sexual sporulation resulting in formation of a cellular spore	113									17	4.80E-04
44085	cellular component biogenesis	960			184	3.19E-62			84	1.38E-29	114	5.39E-30
45943	positive regulation of transcription from RNA polymerase I promoter	12			6	9.43E-03						
46483	heterocycle metabolic process	274									26	6.20E-03
48468	cell development	108									17	2.50E-04
48646	anatomical structure formation involved in morphogenesis	133									17	4.68E-03
51168	nuclear export	121			31	5.33E-11						
51169	nuclear transport	161			34	1.35E-09						
51656	establishment of organelle localization	90			18	5.50E-04						
70590	spore wall biogenesis	43									12	1.77E-05
70591	ascospore wall biogenesis	43									12	1.77E-05
70726	cell wall assembly	44									12	2.35E-05
70925	organelle assembly	79			31	5.57E-17			10	3.82E-03	18	2.45E-07
71035	nuclear polyadenylation-dependent rRNA catabolic	16			7	5.31E-03						
71826	ribonucleoprotein complex subunit organization	134			35	6.41E-13			14	6.60E-04	21	1.17E-05
71840	cellular component organization or biogenesis	2007			200	4.50E-22			91	1.60E-11	124	1.03E-07
71940	fungal-type cell wall assembly	43									12	1.77E-05
90304	nucleic acid metabolic process	1510			186	3.45E-32			84	1.31E-15	91	4.20E-04
90305	nucleic acid phosphodiester bond hydrolysis	122			47	2.57E-26			20	8.94E-10	27	2.00E-11

Table 16: GO enrichment for the different groups of upregulated genes identified in the transcriptome profile of *S. cerevisiae* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

GOID	GO term	Genes in term	Ala		Leu		Lys		Thr		Tyr	
			Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value	Genes	P-Value
5975	carbohydrate metabolic process	305			50	2.16E-05					40	7.06E-06
5977	glycogen metabolic process	32			15	9.44E-07			9	9.62E-03	11	1.30E-04
5991	trehalose metabolic process	11			8	7.76E-05			6	3.17E-03	8	3.20E-06
5996	monosaccharide metabolic process	110			23	2.28E-03						
6007	glucose catabolic process	40					6	4.21E-03				
6066	alcohol metabolic process	208			35	1.55E-03						
6091	generation of precursor metabolites and energy	162			45	6.62E-13			30	1.06E-07	31	3.14E-08
6112	energy reserve metabolic process	34			16	2.42E-07			10	2.04E-03	11	2.60E-04
6119	oxidative phosphorylation	29			13	2.42E-05			11	3.32E-05	12	2.92E-06
6121	mitochondrial electron transport, succinate to ubiquinone	5			5	1.54E-03						
6826	iron ion transport	23							9	4.20E-04		
8643	carbohydrate transport	41					6	4.88E-03				
9056	catabolic process	675			87	2.85E-05			59	3.58E-03	60	2.75E-03
15749	monosaccharide transport	24					6	1.80E-04	8	7.20E-03		
15980	energy derivation by oxidation of organic compounds	138			40	5.87E-12			28	4.77E-08	27	3.49E-07
16052	carbohydrate catabolic process	103			23	6.80E-04	9	3.08E-03	17	5.62E-03	20	7.38E-05
16137	glycoside metabolic process	24							8	7.20E-03	8	8.07E-03
33212	iron assimilation	11							9	4.72E-08		
33214	iron assimilation by chelation and transport	9							7	1.34E-05		
34637	cellular carbohydrate biosynthetic process	91			20	4.55E-03						
42026	protein refolding	17									7	5.73E-03
42775	mitochondrial ATP synthesis coupled electron transport	28			12	1.50E-04			11	2.15E-05	11	2.51E-05
44042	glucan metabolic process	47			16	6.77E-05					13	1.80E-04
44281	small molecule metabolic process	731			96	1.74E-06					63	4.24E-03
44282	small molecule catabolic process	128			30	4.34E-06	11	3.70E-04	20	2.18E-03	20	2.73E-03
45333	cellular respiration	91			21	1.15E-03			16	4.48E-03		
46164	alcohol catabolic process	58					7	3.58E-03				
46365	monosaccharide catabolic process	51					7	1.50E-03				
55065	metal ion homeostasis	61							13	3.69E-03		
55114	oxidation-reduction process	171			46	1.19E-12	13	1.60E-04	34	7.25E-10	34	1.10E-09

Table 17: Overlap of 3 or more deregulated genes in *C.albicans* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

			Fold				
Process	***C. albicans* orf19 Name**	***C. albicans* Common Name**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
cellular protein modification process	orf19.3225	CWH43	-1.22	-1.36			-1.14
cellular protein modification process	orf19.3267		1.22	-1.25			-1.29
DNA metabolic process	orf19.2469	RAD10	1.75	1.43	1.28		1.29
pathogenesis	orf19.5399	IFF11	1.63	8.34		1.78	
response to chemical stimulus /drug	orf19.2395	IME2	5.28	2.35			1.64
response to chemical stimulus /drug	orf19.5621		-2.04	-2.06			-1.88
RNA metabolic process	orf19.265		222.56	8.65		74.54	
RNA metabolic process	orf19.332		-1.36	1.54			1.44
RNA metabolic process	orf19.5486.1	SMD2	-1.47	-2.38			-1.53
RNA metabolic process	orf19.5964.2	RPL35	-1.32	-1.43			-1.68
transport	orf19.1142		1.55	1.49		1.39	
transport	orf19.1357	FCY21	1.41	1.73			1.35
transport	orf19.1867		20.19	6.13		12.15	
transport	orf19.2848		3.41	1.70			1.70
transport	orf19.4036	APM1	-1.58	-1.34			-1.51
transport	orf19.4599	PHO89	1.49	1.63			1.39
transport	orf19.6435		1.24	-1.35			-1.38
transport	orf19.7205	DUR7	1.26	1.36		1.26	
transport	orf19.7331	FCY24	2.65	3.02			2.20
transport	orf19.7637	YHB4	2.31	2.74	1.31		
vitamin metabolic process	orf19.277	THI6	1.67	1.46			1.38
biological process unknown	orf19.1377	IPK2	-1.18	-1.60			-1.19
biological process unknown	orf19.1799		1.34	2.01			1.65
biological process unknown	orf19.441	RPT1		1.37		1.17	1.35
biological process unknown	orf19.6813		-1.16	-1.94			-1.46
biological process unknown	orf19.1169		-1.86	-2.32	-1.56		
biological process unknown	orf19.1864			1.25	1.10		1.22
biological process unknown	orf19.3279	HYR4	1.51	1.61			1.68
biological process unknown	orf19.720	GST3	2.78	1.92			2.15
biological process unknown	orf19.1235	HOM3	2.07	1.62	1.40	1.29	
biological process unknown	orf19.1639		2.15	2.30		1.81	1.71
biological process unknown	orf19.1375	LEU42	5.36	2.17	1.51		
biological process unknown	orf19.1847	ARO10	1.79	1.91			2.59
biological process unknown	orf19.2065		1.47	1.40	1.25		
biological process unknown	orf19.2251	AAH1	1.27	5.75	1.16		
biological process unknown	orf19.2989	GOR1	1.93	2.18			1.80
biological process unknown	orf19.6138		3.72	1.87			1.43
biological process unknown	orf19.742	ALD6	2.08	4.56	2.02		
biological process unknown	orf19.923	THR1	1.26	1.61			1.35

Table 18: Overlap of 3 or more deregulated genes in *C.albicans* strains grown at 37ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

			Fold				
Process	***C. albicans* orf19 Name**	***C. albicans* Common Name**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
cell wall organization	orf19.4581		-1.71		-1.69	-2.12	
interspecies interaction between organisms	orf19.3380	HWP2		-1.38	-2.37	-2.16	
protein catabolic process	orf19.6630		1.38		-1.29		1.39
regulation of biological process	orf19.4251	ZCF22	-1.58		-1.32		-1.81
regulation of biological process	orf19.6810		-1.29	-1.38			-1.30
RNA metabolic process	orf19.4033	PRP22	1.10	1.14			1.09
transport	orf19.3746	IFC1	-2.32			-1.77	1.37
transport	orf19.4228		-1.11		-1.09		-1.06
transport	orf19.4811		1.23		-1.41		1.35
transport	orf19.4961	STP2	-1.48	-1.63			-1.47
transport	orf19.508	QDR1			-1.51	-1.71	-1.51
transport	orf19.7331	FCY24	-1.75	-1.62	-1.43		
vitamin metabolic process	orf19.889	THI20		-1.39	-1.21	-1.29	1.37
biological process unknown	orf19.3879			-1.35	-1.26		-1.21
biological process unknown	orf19.4263		1.44	1.33		1.62	
biological process unknown	orf19.94		1.95		2.14	2.73	
biological process unknown	orf19.4114	FAA2-1	1.28	1.43		1.59	
biological process unknown	orf19.597				-1.57	-1.78	1.21
biological process unknown	orf19.6305			-1.56	1.34		-1.21

Table 19: Overlap of 4 or more deregulated genes in *S. cerevisiae* strains grown at 30ºC and misincorporating serine in alanine, leucine, lysine, threonine or tyrosine codons.

			Fold				
Process	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
cell wall organization or biogenesis	YBR005w	RCR1	1.27	1.30		1.34	1.44
cellular amino acid metabolic process	YFL030w	AGX1	1.65	4.17	1.68	2.33	2.93
cellular amino acid metabolic process	YLR089c	ALT1	1.35	1.73		1.26	1.63
cofactor metabolic process	YNL200c	YNL200c		2.43	1.35	1.72	1.63
cofactor metabolic process	YLR201c	COQ9		1.49	1.35	1.48	1.41
cytoplasmic translation	YOR091w	TMA46		-1.68	-1.29	-1.88	-1.53
cytoplasmic translation	YGR173w	RBG2	-1.34	-1.74		-1.62	-1.45
histone modification/ chromatin organization	YJR119c	JHD2		1.45	1.28	1.40	1.46
lipid metabolic process	YGR110w	CLD1	2.09	2.80	1.54	1.83	2.57
lipid metabolic process	YPL206c	PGC1	1.33	1.41	1.48	1.61	1.49
lipid metabolic process	YLR193c	UPS1	1.15	1.24	1.22		1.14
lipid metabolic process	YPL110c	GDE1	-1.13	1.57	-1.20	1.27	
membrane fusion	YHR138c	YHR138c	1.51	2.30		1.62	1.81
mitochondrion organization	YAL048c	GEM1	1.23	1.28		1.31	1.19
mitochondrion organization	YDR379c-a	YDR379c-a	1.48	2.81	1.64	2.62	2.55
peptidyl-amino acid modification	YIL064w	SEE1	-1.33	-2.27		-1.63	-1.64
peptidyl-amino acid modification	YBR271w	YBR271w	-1.27	-1.86		-1.73	-1.48
peptidyl-amino acid modification	YLL022c	HIF1		-1.19	1.11	-1.15	-1.18
protein complex biogenesis	YLR199c	PBA1	1.19	1.36	1.16	1.22	1.31
protein maturation	YER078c	ICP55	-1.15		-1.13	-1.25	-1.15
protein phosphorylation	YPL150w	YPL150w	1.11	1.28		1.13	1.23
proteolysis involved in cellular protein catabolic process	YDR161w	YDR161w	-1.45	-1.94		-1.91	-1.72
response to chemical stimulus	YHR029c	YHI9	1.56	1.86	1.21		1.82
ribosomal large subunit biogenesis	YKL082c	RRP14		-2.14	-2.36	-2.16	-1.96
ribosomal large subunit biogenesis	YKL014c	URB1	-1.24	-2.03	-1.15	-1.53	-1.57
RNA modification/ tRNA processing	YOL125w	TRM13	-1.26	-2.45	-1.24	-1.70	-1.33
sporulation	YNL194c	YNL194c	2.31	8.92	2.67	4.09	3.73
sporulation	YDL222c	FMP45		4.25	1.69	2.62	2.66
transcription from RNA polymerase II promoter	YOR338w	YOR338w	2.09	2.09	1.55	1.89	1.98
transcription from RNA polymerase II promoter	YLL029w	FRA1	1.28	1.43		1.33	1.41
transcription from RNA polymerase II promoter	YOR172w	YRM1	1.16	1.27		1.13	1.21
transcription from RNA polymerase II promoter	YDR223w	CRF1	1.38	2.10		1.38	1.82
transmembrane transport	YBR241c	YBR241c	1.42	2.44	1.60	1.73	1.88
transmembrane transport	YDL199c	YDL199c	1.65	2.25		1.71	2.23
transmembrane transport	YBR043c	QDR3	1.30	1.44		1.22	1.39
biological process unknown	YDL023c	YDL023c	1.63	1.88	1.50	1.67	1.56
biological process unknown	YIR036c	IRC24	1.20	1.38	1.38	1.51	1.33
biological process unknown	YLR327c	TMA10	2.13	11.66	-1.37	3.91	6.96
biological process unknown	YDR491c	YDR491c	2.48	2.27	2.30	1.67	1.57
biological process unknown	YBR139w	YBR139w	1.34	1.86	1.48	1.91	2.03
biological process unknown	YBR116c	YBR116c	1.95	7.39	1.85	1.51	2.18
biological process unknown	YFR045w	YFR045w	1.25	1.45	1.34	1.54	1.35
biological process unknown	YGL118c	YGL118c	2.09	2.37	1.92	2.20	1.79
biological process unknown	YMR090w	YMR090w	1.23	2.42	1.80	1.96	2.03
biological process unknown	YBL048w	YBL048w	2.66	9.82	2.34	3.73	4.28
biological process unknown	YGR043c	NQM1	1.47	6.72	1.68	2.15	2.63
biological process unknown	YBR285w	YBR285w	1.75	10.51	1.97	8.98	2.35
biological process unknown	YMR173w-a	YMR173w-a	1.61	2.84	1.55	2.01	2.28
biological process unknown	YGL146c	RRT6	1.16	2.29	1.16	1.89	1.46
biological process unknown	YDR070c	FMP16	2.61	17.07	3.92	3.87	5.26
biological process unknown	YGR066c	YGR066c	2.31	3.03	1.74	3.21	2.46
biological process unknown	YJR107w	YJR107w	1.31	1.29	1.28	1.40	1.42
biological process unknown	YDR034w-b	YDR034w-b	2.72	7.21	2.13	6.71	2.72
biological process unknown	YOR289w	YOR289w	1.67	4.14	1.44	2.42	3.62
biological process unknown	YNL208w	YNL208w	1.35	1.80	1.28	1.47	1.32
biological process unknown	YHR112c	YHR112c	1.35	1.72	1.20	1.30	1.30
biological process unknown	YMR118c	YMR118c	2.10	9.42	1.65	2.35	1.98
biological process unknown	YDR230w	YDR230w	1.34	1.63	1.38	1.40	1.32
biological process unknown	YLR143w	YLR143w	-1.23	-1.35	-1.19	-1.40	-1.19
biological process unknown	YLR252w	YLR252w	1.39	3.09	1.47	1.99	2.32
biological process unknown	YCR076c	YCR076c	1.21	1.58	1.24	1.34	1.39
biological process unknown	YER158c	YER158c	1.73	2.33	2.01	1.68	1.57
biological process unknown	YLL020c	YLL020c	1.52	1.98		1.67	1.95
biological process unknown	YPL264c	YPL264c	1.47	1.60		1.43	1.69

			Fold				
Process	***S. cerevisiae* Systematic Name**	***S. cerevisiae* Common Name**	**Ala**	**Leu**	**Lys**	**Thr**	**Tyr**
biological process unknown	YJL163c	YJL163c	1.41	2.84		1.94	1.76
biological process unknown	YMR110c	HFD1	1.15	1.66		1.48	1.48
biological process unknown	YER039c-a	YER039c-a	1.55	1.77		1.99	2.14
biological process unknown	YIL089w	YIL089w	1.23	1.29	1.22	1.47	
biological process unknown	YEL067c	YEL067c	1.20		1.28	1.33	1.22
biological process unknown	YIL055c	YIL055c	1.20	1.67		1.42	1.62
biological process unknown	YNL115c	YNL115c		1.75	1.30	1.50	1.73
biological process unknown	YCR007c	YCR007c	1.21	1.51		1.25	1.27
biological process unknown	YOR309c	YOR309c	-1.26	-2.00		-1.74	-1.73
biological process unknown	YKR011c	YKR011c	1.46	1.90		1.85	1.68
biological process unknown	YML116w-a	YML116w-a		1.36	1.10	1.16	1.25
biological process unknown	YHR209w	CRG1	-1.18	1.56	-1.07	1.86	
biological process unknown	YNL274c	GOR1		2.79	1.37	2.34	2.20
biological process unknown	YDR061w	YDR061w	1.06	1.22		1.20	1.14
biological process unknown	YPR117w	YPR117w		1.37	1.25	1.34	1.38
biological process unknown	YML089c	YML089c	1.42	3.04		2.92	1.64
biological process unknown	YOR228c	YOR228c		1.52	1.28	1.37	1.39
biological process unknown	YJL144w	YJL144w		2.69	-2.02	1.75	1.95
biological process unknown	YJR111c	YJR111c	1.25	1.23	1.23		1.23
biological process unknown	YBR230c	OM14		2.51	1.47	1.78	2.06
biological process unknown	YGR201c	YGR201c	1.49	2.65	1.52		1.56
biological process unknown	YIL087c	AIM19	1.34	2.35	1.48		1.79
biological process unknown	YMR052c-a	YMR052c-a	1.29	2.33		1.79	1.85
biological process unknown	YMR191w	SPG5	1.34		1.33	1.51	1.25
biological process unknown	YNL171c	YNL171c	1.70		1.61	1.85	1.58
biological process unknown	YMR181c	YMR181c	1.84	3.00		2.13	2.37
biological process unknown	YNL092w	YNL092w	1.26	1.85		1.66	1.67
biological process unknown	YLR164w	YLR164w	1.31	1.70	1.30		1.33
biological process unknown	YLR311c	YLR311c		7.09	1.65	4.92	4.45
biological process unknown	YLR149c	YLR149c	1.44	4.74		2.67	3.60
biological process unknown	YPR151c	SUE1	1.57	6.07		2.40	2.11
biological process unknown	YBR137w	YBR137w	1.08	1.22		1.15	1.19
biological process unknown	YOR343c	YOR343c	1.27	3.48		1.83	1.38
biological process unknown	YHR080c	YHR080c	1.05	1.66	1.12		1.49
biological process unknown	YMR148w	OSW5	1.23	1.30	1.39		1.34
biological process unknown	YOR139c	YOR139c		1.71	1.38	1.74	1.68
biological process unknown	YOR152c	YOR152c		2.20	1.54	1.80	1.74
biological process unknown	YNR062c	YNR062c	1.27	1.55	1.40		1.25
biological process unknown	YMR196w	YMR196w		3.38	1.96	2.13	2.56
biological process unknown	YGL052w	YGL052w	1.25	2.04		1.83	1.88
biological process unknown	YBR281c	DUG2	-1.14	-1.39	-1.16		-1.14
biological process unknown	YOR220w	RCN2	1.52	1.78		1.62	1.61
biological process unknown	YPL250c	ICY2	1.95	2.29		2.05	2.03
biological process unknown	YNR068c	YNR068c	1.47	2.34		1.29	2.20
biological process unknown	YBR094w	PBY1	-1.20	-1.16	-1.16		-1.12
biological process unknown	YFR020w	YFR020w	-1.20	-1.18		-1.27	-1.17
biological process unknown	YPL222w	FMP40	1.30	1.94		1.53	1.42
biological process unknown	YDL110c	TMA17	1.32	2.41		1.69	1.96
biological process unknown	YIL086c	YIL086c	-1.18	-1.54		-1.20	-1.24
biological process unknown	YLR312c	YLR312c	1.41	4.88		3.21	2.66
biological process unknown	YOL014w	YOL014w	-1.38	-2.93	-1.61		-2.67
biological process unknown	YMR310c	YMR310c	-1.16	-1.58	1.14		-1.51
biological process unknown	YGR236c	SPG1	1.42	3.36		1.65	1.51
biological process unknown	YML009w-b	YML009w-b	-1.35	-1.47		-1.36	-1.40
biological process unknown	YLR407w	YLR407w	1.16	1.20	1.14		1.29
biological process unknown	YBR204c	YBR204c		1.26	-1.12	1.26	1.23
biological process unknown	YBR033w	EDS1	1.73	3.17		2.11	1.93
biological process unknown	YMR086w	YMR086w	1.19	1.42	1.25		1.28
biological process unknown	YNL010w	YNL010w	-1.06	-1.42		-1.20	-1.31
biological process unknown	YDL085c-a	YDL085c-a	-1.25	-1.29		-1.19	-1.15
biological process unknown	YMR107w	SPG4		11.61	1.35	1.89	4.82
biological process unknown	YPL191c	YPL191c	1.19	1.46		1.32	1.51
biological process unknown	YOL098c	YOL098c	-1.07	-1.23		-1.17	-1.10
biological process unknown	YHR087w	RTC3		4.51	2.02	2.99	4.95

Table 20: Transcription Factors from the overlap genes described in table 17, 18 and 19.

Transcription factor			C. albicans (%)			S.cerevisiae (%)	
S.cerevisiae	*C.albicans orthologs*		30ºC	37ºC	Random	30ºC	Random
	Orf.	Gene Name					
Abf1p			-	-	6.0	46.1	60.2
Aca1p	ORF19.6102		2.6	5.3	1.0	5.5	4.1
Ace2p	ORF19.6124	ACE2	20.5	15.8	27.0	25.8	24.5
Adr1p	ORF19.2961, ORF19.5975, ORF19.2752	ADR1 (ORF19.2752)	59.0	52.6	64.0	59.4	56.1
Aft1p	ORF19.2272		7.7	5.3	13.0	13.3	16.3
Aft2p	ORF19.2272		7.7	5.3	13.0	17.2	16.3
Arg81p	ORF19.4766, ORF19.2748		-	-	-	18.0	13.3
Ash1p	ORF19.5343	ASH1	100.0	100.0	96.0	99.2	100.0
Azf1p	ORF19.173		69.2	63.2	74.0	52.3	51.0
Bas1p	ORF19.3809		38.5	31.6	34.0	25.8	25.5
Cad1p	ORF19.1623	CAP1	28.2	21.1	25.0	14.8	22.4
Cat8p	ORF19.5097, ORF19.2280, ORF19.6817	CAT8 (ORF19.5097),FCR1 (ORF19.6817)	-	10.5	2.0	5.5	2.0
Cbf1p	ORF19.2876	CBF1	10.3	15.8	18.0	10.2	8.2
Cha4p	ORF19.5729	FGR17	-	-	-	18.8	11.2
Cin5p			28.2	21.1	25.0	14.8	22.4
Crz1p	ORF19.2356, ORF19.4972, ORF19.7359	CRZ2(ORF19.2356), CRZ1(ORF19.7359)	7.7	10.5	15.0	15.6	8.2
Cst6p			2.6	5.3	1.0	5.5	4.1
Cup2p	ORF19.5001	CUP2	43.6	52.6	48.0	55.5	52.0
Dal81p	ORF19.3252		2.6	-	1.0	3.9	1.0
Dal82p			2.6	-	1.0	1.6	-
Ecm22p	ORF19.2077, ORF19.2623, ORF19.391	ECM22 (ORF19.2623), UPC2 (ORF19.391)	2.6	5.3	9.0	10.9	12.2
Fkh1p	ORF19.5389	FKH2	100.0	100.0	98.0	98.4	98.0
Fkh2p	ORF19.5389	FKH2	100.0	100.0	98.0	98.4	98.0
Flo8p	ORF19.1093	FLO8	-	-	-	0.8	-
Gal4p	ORF19.5338	GAL4	-	-	-	5.5	6.1
Gat1p	ORF19.1275, ORF19.1150	GAT1 (ORF19.1275)	48.7	78.9	30.0	45.3	40.8
Gcn4p	ORF19.1358	GCN4	61.5	47.4	61.0	53.1	61.2
Gcr1p			84.6	89.5	91.0	97.7	99.0
Gis1p	ORF19.2730, ORF19.2743		51.3	52.6	57.0	68.8	54.1
Gln3p	ORF19.3912, ORF19.1150	GLN3(ORF19.3912)	48.7	78.9	30.0	57.8	58.2
Gsm1p			-	15.8	4.0	21.9	14.3
Gzf3p	ORF19.2842	GZF3	48.7	78.9	30.0	45.3	40.8
Hac1p	ORF19.2432	HAC1	46.2	68.4	59.0	63.3	68.4
Hap1p	ORF19.1255, ORF19.5133, ORF19.7372	MRR1(ORF19.7372)	-	-	-	7.0	4.1
Hap2p	ORF19.1228	HAP2	46.2	42.1	33.0	14.1	11.2
Hap3p	ORF19.4647, ORF19.517	HAP3 (ORF19.4647), HAP31(ORF19.517)	46.2	42.1	33.0	14.1	11.2
Hap4p	ORF19.681	HAP43	46.2	42.1	33.0	14.1	11.2
Hap5p	ORF19.1973	HAP5	46.2	42.1	33.0	14.1	11.2
Hcm1p	ORF19.4853	HCM1	12.8	10.5	11.0	1.6	3.1
Hsf1p	ORF19.4775	CTA8	41.0	68.4	55.0	55.5	55.1
Ime1p			10.3	5.3	14.0	16.4	15.3
Ino2p			-	-	5.0	0.8	4.1
Ino4p	ORF19.837.1	INO4	-	-	5.0	0.8	5.1
Kar4p	ORF19.3736	KAR4	5.1	-	-	-	1.0
Leu3p	ORF19.4225	LEU3	-	-	1.0	6.2	5.1
Lys14p	ORF19.4778, ORF19.4776, ORF19.5548, ORF19.6781, ORF19.6888, ORF19.3405		5.1	-	6.0	2.3	3.1
Mac1p	ORF19.7068	MAC1	7.7	5.3	7.0	9.4	11.2
Mbp1p	ORF19.5855		2.6	10.5	9.0	9.4	5.1
Mcm1p	ORF19.7025	MCM1	38.5	31.6	39.0	60.2	59.2
Met31p	ORF19.1757		12.8	-	7.0	8.6	7.1
Met32p	ORF19.1757		12.8	-	7.0	8.6	7.1
Met4p	ORF19.5312, ORF19.7046	MET28 (ORF19.7046)	5.1	-	13.0	7.0	4.1
Mot3p			100.0	100.0	98.0	100.0	100.0
Msn2p	ORF19.7150	NRG1	51.3	52.6	57.0	60.9	53.1
Msn4p	ORF19.4752	MSN4	51.3	52.6	57.0	60.9	53.1
Mss11p			-	-	-	0.8	-
Ndt80p	ORF19.513, ORF19.2119	NDT80 (ORF19.2119)	17.9	15.8	11.0	18.8	15.3
Nrg1p	ORF19.7150	NRG1	51.3	52.6	46.0	82.0	81.6
Oaf1p	ORF19.7374	CTA4	-	-	-	3.1	3.1
Pdr1p	ORF19.4573, ORF19.6182		2.6	5.3	-	6.2	4.1
Pdr3p			2.6	5.3	-	6.2	4.1
Pdr8p			-	5.3	-	5.5	7.1
Pho4p	ORF19.1253		30.8	26.3	33.0	41.4	30.6
Pip2p			-	-	-	3.1	3.1
Ppr1p	ORF19.3986		-	-	-	5.5	6.1
Put3p	ORF19.1773, ORF19.4568, ORF19.6203	RAP1(ORF19.1773), ZCF25 (ORF19.4568)	-	-	-	7.8	4.1
Rap1p			-	-	-	-	2.0
Reb1p	ORF19.5722		-	-	-	5.5	3.1
Rfx1p	ORF19.3865, ORF19.4590		-	-	-	1.6	-
Rgt1p	ORF19.2747	RGT1	51.3	52.6	44.0	74.2	69.4
Rim101p	ORF19.3434, ORF19.7247	RIM101 (ORF19.7247)	7.7	26.3	6.0	-	-
Rlm1p	ORF19.4662	RLM1	20.5	15.8	20.0	17.2	16.3
Rme1p	ORF19.4438	RME1	-	-	-	0.8	-
Rox1p	ORF19.2823	RFG1	-	5.3	-	-	2.0
Rph1p			-	-	-	60.9	53.1

Transcription factor			C. albicans (%)			S.cerevisiae (%)	
S.cerevisiae	*C.albicans orthologs*		30ºC	37ºC	Random	30ºC	Random
	Orf.	Gene Name					
Rpn4p	ORF19.1069, ORF19.2054	RPN4 (ORF19.1069), FGR15 (ORF19.2054)	2.6	-	2.0	2.3	1.0
Rtg1p	ORF19.4722		76.9	89.5	77.0	89.1	85.7
Rtg3p	ORF19.2315		76.9	89.5	77.0	89.1	85.7
Sfl1p	ORF19.3969, ORF19.454	SFL1 (ORF19.454)	-	5.3	6.0	3.1	5.1
Sip4p			-	-	-	5.5	2.0
Skn7p	ORF19.971	SKN7	5.1	-	9.0	16.4	9.2
Sko1p	ORF19.1032	SKO1	7.7	-	3.0	7.8	7.1
Stb5p	ORF19.3308, ORF19.7583	N/A	89.7	84.2	69.0	99.2	96.9
Ste12p	ORF19.4433	CPH1	28.2	42.1	26.0	24.2	20.4
Stp1p	ORF19.4961	STP2	-	-	-	11.7	11.2
Stp2p	ORF19.4961, ORF19.5917	STP2 (ORF19.4961), STP3 (ORF19.5917)	10.3	10.5	10.0	26.6	20.4
Sum1p			17.9	15.8	11.0	18.8	15.3
Swi4p	ORF19.4545	SWI4	7.7	-	15.0	16.4	25.5
Swi5p	ORF19.2612		20.5	15.8	27.0	25.8	24.5
Tec1p	ORF19.5908	TEC1	79.5	78.9	77.0	71.1	68.4
Uga3p	ORF19.6038, ORF19.7317, ORF19.7317		5.1	-	1.0	0.8	2.0
Ume6p	ORF19.2745,ORF19.3127, ORF19.1822	CZF1, UME6 (ORF19.1822)	-	-	-	2.3	3.1
Upc2p	ORF19.3305		2.6	5.3	9.0	14.1	16.3
War1p	ORF19.1035, ORF19.6985	WAR1(ORF19.1035), TEA1(ORF19.6985)	-	-	-	3.9	9.2
Xbp1p	ORF19.5210		53.8	47.4	39.0	60.2	60.2
Yap1p	ORF19.1623	CAP1	76.9	73.7	65.0	60.2	60.2
Yap3p	ORF19.3193	FCR3	28.2	21.1	25.0	20.3	29.6
Yap5p			28.2	21.1	25.0	14.8	22.4
Yrr1p			25.6	5.3	21.0	24.2	29.6
Zap1p	ORF19.3794	CSR1	-	-	-	1.6	3.1